

STIC-Biotech/Ch mLib

76159

From: Rao, Manjunath N.  
Sent: Friday, September 20, 2002 10:12 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 09/836613

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10C 01  
Phone: 306-5681

Date: 9-20-02

RECEIVED  
SEP 20 2002  
STIC/Biotech Division  
(STIC)

Please search the following as soon as possible for application with serial number **09/836,613**

**SEQ ID NO:2**, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If y u have any questions please call me at the above phone number.

Thanks

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

Manjunath N. Rao, Ph.D.  
Biot chnology Patent Examiner  
Art Unit 1652, Room 10A11  
Crystal Mall 1, USPTO.

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 9/21/02  
Date Completed: 9/21/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: ✓  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 12:23:45 ; Search time 34.74 Seconds  
(without alignments)  
2375.585 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAAGVLLLAGAGG.....VDLAKKIFLKYYPGWAGSW 743

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1980.DAT.\*
- 2: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1981.DAT.\*
- 3: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1982.DAT.\*
- 4: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1983.DAT.\*
- 5: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1984.DAT.\*
- 6: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1985.DAT.\*
- 7: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1986.DAT.\*
- 8: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1987.DAT.\*
- 9: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1988.DAT.\*
- 10: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.\*
- 11: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1989.DAT.\*
- 12: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1990.DAT.\*
- 13: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1991.DAT.\*
- 14: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1992.DAT.\*
- 15: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.\*
- 16: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1994.DAT.\*
- 17: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1995.DAT.\*
- 18: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1996.DAT.\*
- 19: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1997.DAT.\*
- 20: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1998.DAT.\*
- 21: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA1999.DAT.\*
- 22: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA2000.DAT.\*
- 23: /SIDSL1/gcgdata/hold-geneseq/geneqseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3939	100.0	743	18 AAW18017	Human alpha-N-acet
2	1407	35.7	690	22 ABB68119	Drosophila melanog
3	125.5	3.2	497	22 AAE05171	Human drug metabol
4	121	3.1	4342	22 AAU33611	Pseudomonas aerugi
5	121	3.1	4472	18 AAU22601	Tyrlactone synthase
6	120	3.0	3729	18 AAU22603	Tyrlactone synthase
7	118.5	3.0	871	22 AAB30570	A splice variant o
8	118.5	3.0	1351	22 AAB30567	Amino acid sequenc
9	118.5	3.0	1610	22 AAB30568	A full length huma
10	118.5	3.0	1618	22 AAB85504	Human protein kina
11	118.5	3.0	2596	22 AAB30569	A splice variant o

12	117.5	3.0	497	16 AAR72376	Human auxillary cy
13	117.5	3.0	497	16 AAR72377	Human auxillary cy
14	117.5	3.0	497	17 AAR93183	Human cytochrome p
15	117.5	3.0	497	17 AAR93184	Human cytochrome p
16	117.5	3.0	497	17 AAR81462	Human derived cyto
17	117.5	3.0	497	19 AAW44869	Cytochrome P450Iid
18	113	2.9	7257	21 AAV58576	Sorangium cellulos
19	111.5	2.8	484	21 AAG45186	Arabidopsis thalia
20	111.5	2.8	497	16 AAR72375	Human auxillary cy
21	111.5	2.8	497	17 AAR93182	Human auxillary cy
22	111.5	2.8	497	17 AAR93185	Human cytochrome p
23	111.5	2.8	497	17 AAR93185	Human cytochrome p
24	109.5	2.7	1377	22 AAU34791	E. coli cellular p
25	109.5	2.8	1377	22 AAG98997	E. coli growth and
26	107.5	2.7	1621	19 AAG60182	Arabidopsis thalia
27	106	2.7	308	22 AAG92974	C glutamicum prote
28	106	2.7	1005	22 ABG11447	Novel human diagno
29	105	2.7	1203	20 AAW81572	Novel human diagno
30	105	2.7	4150	21 AAY92707	S. antibioticus 8,
31	104.5	2.7	1622	19 AAW60163	Arabidopsis thalia
32	104	2.6	581	21 AAY84322	A human cardiovasc
33	104	2.6	581	22 AAB65701	Novel protein kina
34	104	2.6	588	21 AAY58161	Adeno associated v
35	104	2.6	654	22 AAU29036	Human PRO polypept
36	104	2.6	654	22 AAE06593	Human protein havi
37	104	2.6	654	22 AAB80243	Human PRO282 prote
38	104	2.6	668	22 AAB67462	Amino acid sequenc
39	104	2.6	677	17 AAR99539	Soluble starch syn
40	104	2.6	724	21 AAY58160	Adeno associated v
41	103	2.6	1188	17 AAR94622	HIV-1 gp160-NV5 p5
42	102	2.6	1646	22 AAR79494	Human protein SEQ
43	102	2.6	2308	15 AAR57902	Human RPTP-beta.
44	101.5	2.6	5069	19 AAW52846	A. mediterranei ri
45	101	2.6	1231	22 ABG17594	Novel human diagno

#### ALIGNMENTS

RESULT 1

AAW18017

ID AAW18017 standard; Protein; 743 AA.

XX AAW18017;

AC AAW18017;

DT 20-AUG-1997 (first entry)

XX Human alpha-N-acetylglucosaminidase.

DE Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;

KW gene therapy; enzyme replacement therapy; diagnosis.

KW Homo sapiens.

OS Homo sapiens.

XX Key

FH Peptide

FT Location/Qualifiers

FT 1..23

FT /label= Sig\_peptide

FT 22..743

FT /label= Mat\_protein

FT 261

FT Modified-site

FT /note= "potential N-glycosylation site"

FT 272

FT Modified-site

FT /note= "potential N-glycosylation site"

FT 435

FT Modified-site

FT /note= "potential N-glycosylation site"

FT 503

FT Modified-site

FT /note= "potential N-glycosylation site"

FT 513

FT Modified-site

FT /note= "potential N-glycosylation site"

FT 526

FT Modified-site

FT /note= "potential N-glycosylation site"

FT 532

FT Modified-site

FT /note= "potential N-glycosylation site"



Query Match		35.7%	Score 1407;	DB 22;	Length 690;
Best Local Similarity		43.3%	Pred. No. 1.4e-118;		
Matches		290;	Conservative 102;	Mismatches 255;	Indels 22; Gaps 11;
QY	74	RVRVGGTGVAAAAGLHRYLRDFCGCHVANGSGQLRLPR--PLPAPVGEITLTPNRYR	131		
Db	6	rillmgwgvsvckalhhkyivlnkdvdwfkmlieptnlqlpnvtieskasp--ily	63		
QY	132	YQNVCTOSYFVWMDWEREDWALNGINLALAWSGOEAIWORYVIALGLTQAEINE	191		
Db	64	hqnvtswsfawgileqrhrldwalmgisitia-pvqealwkvvytdmgirmeide	122		
QY	192	FFTCGPAFLANGMGNLHTWDPGLPPSHHIIKOLYLQHRVLDOMRSGMTPLPAFAGHVE	251		
Db	123	hlagpatgawrgnirgwapitpawryqlllqqeiltacnlgmsvalpafaghvpr	182		
QY	252	AVTRVFPQVNVTKMGSHGNCYSYCSFLAPEDPFPPIIGSLFLRELKEFGTDHIYGA	311		
Db	183	alkrlpestfmevqrwnqfpyccglfveptenlfkelastrflhniitkygshhffc	242		
QY	312	DTNEMOPSPSPSYLAAATTAIYEAAMTAVDTEAVVLLQGWLFQHQFQWGAQIRAVLG	371		
Db	243	dpineleppvakpeymrstaaiyesmrgidpqaialwlgwmfvknp-fwtmdaeafit	301		
QY	372	AVPRGRLVLDLFAESQPYRTASTQGGPFIWCMHNFNGNHLGFGALEAVYGGPEAAR	431		
Db	302	aapgrilvldlqsedfpdyeltrsyfgqpfwcmhmfgtlmgfsgaklinsgleear	361		
QY	432	LFPNSTMVGTMAGPEISQNEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYGVSHPDAG	491		
Db	362	rlpnssilvgitpegigqymysftlrgwsntsi-dldswftfnishrygvvdkderle	420		
QY	492	AAWRLLRSVYNCSG--EACRGHNRSLVRRPSLQMNSTSIWYNRSDVFAWRLLLTSA	550		
Db	421	qawllknsvyfrglqknrg--qyvvttrpsfndqepftwynaasavldawhlllfrail	478		
QY	551	ATS----PAFRYDLDLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELLPAL	606		
Db	479	plnednryelhdvlditrfqisadqlylnrsaytrkrqvsrf----eflsvklklf	534		
QY	607	DE---VLASDSRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWPGNGLDYANKOL	663		
Db	535	dmeilassrnflignwlqqkaapntgqgrnfefnarnqitawpgdgildyackqw	594		
QY	664	AGLVANYTPRWRLFLEALVDSVAQIPFQHQFDKNV-FOLEQAFVLSKORYSPQPRGD	722		
Db	595	sglvdyrrprwrlfledvtvalhagrpfbngtatkikvshelipfsnkdkdyvypvpgn	654		
QY	723	TVDLAKKIF 731			
Db	655	twllsqdif 663			
RESULT 3					
ID	AAE05171	standard; Protein; 497 AA.			
XX	AC	AAE05171;			
XX	DT	12-SEP-2001 (first entry)			
XX	DE	Human drug metabolising enzyme (DME-2) protein.			
XX	KW	Human; drug metabolising enzyme; DME-2; immunosuppressive; gene therapy;			
KW	KW	cytostatic; autoimmune disorder; hepatic tumour; inflammatory disorder; atherosclerosis;			
KW	KW	osteoporosis; eye disorder; Addison's disease; cretinism;			
KW	KW	rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;			
KW	KW	developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;			
KW	KW	thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;			
KW	KW	gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;			
KW	KW	atkinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;			
XX	XX	cell proliferative disorder.			
OS					
XX	XX	Homo sapiens.			
XX	XX	Key			
FT	FT	Peptide			
FT	FT	Protein			
FT	FT	Domain			
FT	FT	Domain			
XX	XX	WO200151638-A2.			
XX	XX	19-JUL-2001.			
XX	XX	12-JAN-2001; 2001WO-US01174.			
XX	XX	14-JAN-2000; 2000US-0176139.			
PR	PR	21-JAN-2000; 2000US-0177443.			
PR	PR	28-JAN-2000; 2000US-0178574.			
XX	XX	(INCY-) INCYTE GENOMICS INC.			
PA	PA	Yang J, Baughn MR, Burford N, Au-Young J, Lu DAM, Reddy R;			
XX	XX	Ring HZ, Hillman JL, Yue H, Azimzai Y, Yao MG, Gandhi AR;			
PI	PI	Nguyen DB, Tang YT, Lal P, Bandman O;			
XX	XX	WPI; 2001-425874/45.			
DR	DR	N-PSDB; AAD09937.			
XX	XX	Drug metabolizing enzymes and encoding polynucleotides, useful for			
PT	PT	diagnosing, treating and/or preventing autoimmune, inflammatory, cell			
PT	PT	proliferative, developmental, endocrine, eye, metabolic, and			
PT	PT	gastrointestinal disorders -			
XX	XX	Claim 1; Page 136-137; 133pp; English.			
PS	PS	The present sequence is human drug metabolising enzyme (DME-2) protein.			
XX	XX	Human DME and its nucleic acid molecule are useful for the diagnosis,			
CC	CC	treatment and prevention of disorders associated with increased or			
CC	CC	decreased expression of DME. Examples of such disorders include,			
CC	CC	autoimmune/inflammatory disorder such as acquired immune deficiency			
CC	CC	syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative			
CC	CC	disorder such as actinic keratosis, atherosclerosis; developmental			
CC	CC	acromegaly, cretinism, thyrotoxicosis; endocrine disorder such as			
CC	CC	diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;			
CC	CC	metabolic disorder; Addison's disease, obesity; gastrointestinal			
CC	CC	nodular hyperplasia, adenomas and carcinomas. DME DNA is useful for			
CC	CC	creating 'knockin' humanised animals (pigs) or transgenic animals (mice			
CC	CC	or rats) to model human disease. DME DNA is also useful for screening			
CC	CC	therapy. DME and its immunogenic fragments are useful for screening			
CC	CC	libraries of compounds in several drug screening assays.			
XX	XX	Sequence 497 AA;			
Query Match					
Best Local Similarity		3.2%;	Score 125.5;	DB 22;	Length 497;
Matches		93;	Conservative 53;	Mismatches 143;	Indels 135; Gaps 21;
QY	345	AVWLL-----QGWLFQHQFQWGAQIRAVLGAVPRGRGLVLDLFAESQPVYTRTAS	396		
Db	14	afillvldlmhrrqrwaarypp---gplplpgl-----gnllhvd-----	50		
QY	397	FOGQPIWCMHNFNGN-----HGLFGALEA-VNGGPEAARL--FPNSTMV	439		
Db	51	fntpycfqdlrrrfqfslqlawtpvvlnglaavrealvthgedtdrppvptqil	110		
QY	440	GTCMAPEGI-----SQNEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYGVSHPD	489		
Db	111	fgfgrsqgvlarypawreqrfsvstlrnlglgk---ksleqwtveaa-----	158		

QY 490 AGAARLLLSVYNGSGEACGRNRSPLVRPSLOMTNTSIWNSDYFEARLLLTSAAPS 549  
Db 159 -----CLcaafanhgrpf--rpnglldkav-----snvl-----as 188  
QY 550 LATSPAFRYD-----LIDLTRQAOELVSLVYEAREASAYLSKELASLLRAGGVLAYE-- 601  
Db 189 ltcgrfeydprflrldlaqegikeesgfrevlnavpvlphlpal--agkvlrfoka 246  
QY 602 LPLALDEVL-----ASDSRFLLSWLEQARAAVSEAEADYEQNSRYQLTLWGPE 652  
Db 247 fitqdeltehrmtwdpaqpprditeafak-kekakgspessfndenlrivv----- 299  
QY 653 GNILDYANKLAGLVANYTPRWRLFLRALVDSVAQGITPFOHQFQDKNVFQLEQAFVLSK 712  
Db 300 gnlf-----lagmvtsttllawallmlhpdvqcrv---qqeidevigvrhpeamadq 350  
QY 713 QRPV 716  
Db 351 ahmp 354  
RESULT 4  
AAU33611  
ID AAU33611 standard; Protein: 4342 AA.  
XX AAU33611;  
AC AAU33611;  
XX  
DT 14-FEB-2002 (first entry)  
XX Pseudomonas aeruginosa cellular proliferation protein #55.  
DE Antisense; prokaryotic cellular proliferation protein;  
XX antibiotic; antibacterial; drug design.  
KW Pseudomonas aeruginosa.  
XX  
OS WO200170955-A2.  
XX  
PN 27-SEP-2001.  
XX  
PD 21-MAR-2001; 2001WO-US09180.  
XX  
PF 21-MAR-2001; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
XX WPI; 2001-611495/70.  
DR N-PSDB; AAS51470.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Example 3; Seq ID NO 5107; 51lpp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4342 AA;

Query Match 3.1%; Score 121; DB 22; Length 4342;  
Best Local Similarity 21.4%; Pred. No. 1.8;  
Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;  
QY 4 VAVAA-----AVGVLLLAGAGG-----AGDEAREAAAVRA-LVARLL 40  
Db 1181 vaicaerspqlivgliaivkagayvpdpdpaperlaymladsgvlllctqahlferl- 1239  
QY 41 GPG-----PAADFSVVERALAAKPGLDYSLGGGAARVRVSGTGVAAAGLHRYLRD 95  
Db 1240 -pgaegvtpicldskldnwpqapglhlh--gdnlayviytsgstgqpkvgv----- 1289  
QY 96 FCGCHVAMSGSOLRLRPLPAVPGCE--LTEATPNRYRYQNVCTOSYSFVWMDWARWERE 153  
Db 1290 ---nthaaala-erlqwmqatytldgdvImqkap-----vfdsvsvw--e 1328  
QY 154 IDWMALINGINLAWSGQ-----EATQWRVYLALGLTQAEINEFFTPGPAFLAWRMG 205  
Db 1329 cfwplvtgcrlylaapgehrdparivelvrgfvtllhtfvpplqlqfidepgvaacgair 1388  
QY 206 NLHTWDGGLPPSNHKLQYLQHRVLDQMRSGFTVLPFAFAGHVPEAVTRVFPQNVTKM 265  
Db 1389 rlfsggealpa-----lrnrviqr-----lpavalhnrvgptet--ainvt-- 1428  
QY 266 GSWGHFNCSYSCSFLAPED---PIFPIIGSLFLRELKEFGTDHIYGAOTFNM----- 317  
Db 1429 ----hwgc-----raedgersplgrplgnvrcvldaef---nllpagvageicgg 1473  
QY 318 -----QPPSEPSYLAATTAAYEAMTAVDTEAVWLLQGLW-----FOHQQPFNG- 362  
Db 1474 lglargylgrpalsaeafvadpfsaagerlyrtgdrrwnadgvieyigrldqqvklrfg 1533  
QY 363 ---PAQIRAVLGAVP---RGRLLVLDLFAESQPV--YTRTASFQGPFFIWCMLHNFNGNH 414  
Db 1534 riepeelgarllaqpgvavviregvagsgqlvgytgvagaeaea-----eqnq 1584  
QY 415 GLFGALEAVNGGPEAARLFPNSTMVTGMAPEGISQNEVVYSMLAELGWRK-----DPVPD 470  
Db 1585 rlrailga-----elpeymvptqlmlragmplgps-gkidtralpepvvqgrehveprte 1638  
QY 471 L-----AAWVTSFAARRYG-----VSHPDAGAAARLLLRSLVYVNCs-- 505  
Db 1639 lqriaaiwseviglprvgirdoffelgghslhatrivstrqacdvplrfalfeasel 1698  
QY 506 -----GEACKGHNRSPLRVRPSLQMNSTSIWNSRDVFEARLLLTSAPLATSPAF 556  
Db 1699 eafceqvraaqagrtshgairridreqvpplaysqgrmfwlql-----epd---spay 1751  
QY 557 RY-----DLDLLTR--QAVQELVSLVYEAREASAYLSKELASLLRA---GGV-LAYELL 603  
Db 1752 nvvgglarlsgpldvvarfeaalqalvq-rhetlrrtspdvqgvprvhgddgimdwqdf 1810  
QY 604 PALD-----EVLASDRSFLLSGWLEQA---RAAAVSEAEADFYEQNSRYQLTLWGPE 652  
Db 1811 saldrdrqhlqlcla-dsahrpfdiespllrivcmkmaerehyllvltlhhiv-----te 1866  
QY 653 GNILDYANKLAGLVANYTYTFRWRLFLRALVDSVAQGITPFOHQFQDKNVFQLE 705  
Db 1867 gwamditarelgalyeaflddr-----esplelpvq--yldysvswqre 1908



XX Tyactone synthase ORF3 protein.  
DE Tyactone synthase gene cluster; tylG gene; multifunctional protein;  
KW polyketide; tyactone synthesis; antibiotic; tylosin.  
KW Streptomyces fradiae.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 1 /note= "encoded by GTC"  
FT Domain 43..465  
FT /note= "ketosynthase domain, KS4"  
FT Domain 588..916  
FT /note= "acyltransferase domain, AT4"  
FT Domain 1178..1356  
FT /note= "ketoreductase domain, KR'4"  
FT Domain 1451..1534  
FT /note= "acyl carrier protein domain, ACP4"  
FT Domain 1556..1980  
FT /note= "ketosynthase domain, KS5"  
FT Domain 2095..2425  
FT /note= "acyltransferase domain, AT5"  
FT Domain 2451..2636  
FT /note= "dehydratase domain, DH2"  
FT Domain 3274..3453  
FT /note= "ketoreductase domain, KR5"  
FT Domain 3548..3631  
FT /note= "acyl carrier protein domain, ACP5"  
XX EP79K55-A2.  
PN 27-AUG-1997.  
XX  
XX 19-FEB-1997; 97EP-0301056.  
XX 22-FEB-1996; 96US-0012078.  
XX (ELIL ) LILLY & CO ELI.  
XX Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;  
PI WPI; 1997-418046/39.  
DR N-PSDB; AAT80413.  
XX  
XX DNA encoding Streptomyces fradiae tyactone synthase domain - for  
XX production of tylosin-related polyketide compounds  
XX  
XX Claim 9; Pages 86-98; 220pp; English.  
XX  
XX AAW22601-W22605 represent proteins encoded by the tyactone synthase  
XX gene cluster of the invention. The gene cluster is also referred to as  
XX the tylG gene, and was isolated from Streptomyces fradiae. These  
XX sequences are multifunctional proteins which direct the synthesis of the  
XX polyketide tyactone, isolated from Streptomyces fradiae. Tyactone is  
XX the basic building block of the antibiotic tylosin. The DNA sequence can  
XX be modified so as to alter the type of carboxylic acids incorporated.  
XX the number of carboxylic acids incorporated and/or the post-condensation  
XX reactions performed, thereby resulting in novel tylosin-related  
XX polyketides.  
XX  
XX Sequence 3729 AA;  
SQ  
Query Match 3.0%; Score 120; DB 18; Length 3729;  
Best Local Similarity 22.4%; Pred. No. 1.7;  
Matches 192; Conservative 72; Mismatches 265; Indels 328; Gaps 47;  
QY 4 VAAVAARV---GVLLLAGAGGAGDEAREAAARVALVRLLG-----PGPAAD 47  
DB 668 vdaavavghsgdaaavagalsle---daavvalragligrlylagamaavplpage 725  
QY 48 FSVSVERALAAKPGDLDTVSLGGGAA-----RVRVRGSGTGVAAAAGLH----- 90

Db 726 ---veaglakwpgveaavngpastvsgdravagvavcqaegvarlipdyashs 781  
QY 91 RYLRDFCG-CHVANGSGLRLPR-PLPAV-----PGLTEATPNRYRYQNVCTQSYSFV 143  
Db 782 rhvedlkgelervlsglrprsvpcstvageqgepv----- 820  
QY 144 WMDWARWER-----EIDMMALNGINL-----ALAWSGQPAIWQRYLA 181  
Db 821 -fdagywfrlnrnrvefsavvggllleeghrfrfievahpvlvhaiegtaeaa--dravha 877  
QY 182 LGLTQAE---INERTGPAFLANGRMGNLHTWDGPPSWHIKOLYL-----QHRVLDDQM 233  
Db 878 tglrrqdsphrltsta-eaahgatl-twdpalpgg-hittlptypfhhhywid-- 932  
QY 234 RSFGMTPLVPAPF-----ACHVPEAVT--RVFPQVNVTKM--GSWGHFNCSYSCS 278  
Db 933 ---tpttptattgspdaqpadaipkyvskwrlrdqdsaltarldgrw----- 978  
QY 279 FLAPEDPIFFIIGSLFLRELK-----EFGTDH-----IYGADTFNEMQPPSS 322  
Db 979 llivpeasadsaevagvareitargetvesltvepgadrsrlrgllvdaterdeagplrg 1038  
QY 323 EPSYLA-----AATTAVYEAMTAVDTEA-VMLQGWLFQHQPFQW 362  
Db 1039 lvsllalagdhagadgarvvpaglaaslaaliqaagdagteaglw----- 1083  
QY 363 PAQIRAVLGAVPGRLLLDLFAESQPV---YTRTASTGQGFQFMCMLHNFEGNHGLFG- 418  
Db 1084 -avtvgavaavpg-----dvpapsqallwgfgrvagie-lphcwggllldiptgpgdsgf 1135  
QY 419 -----ALEAVNGPEAARLPNSTMVGT-GMAPEGISONEVVYSIMAE 461  
Db 1136 rqlaatlagrpaedqvalra--sgayrrflvrasaaggadgwrprg-----tlvv 1184  
QY 462 GWRKDPVPLAAWVTSFAARRY---GVSHP---DAGAAWRLLLRSVYNCSEACRGHNR 514  
Db 1185 gdtavagplvrlwlngarrvtlsglsglplbeeladvaarvtv-----apcdpadr 1236  
QY 515 SPLVRRPSLOMNTSIWYNSDVFEAWRLLLTSAPSLATSPAFRYDLDDLTROA----- 567  
Db 1237 palrtllaegaptav-----lvappavppcp-----laemtaealatale 1277  
QY 568 ----VQELVSLYE-----EARSAYLSKEIASLL 592  
Db 1278 ktglvrlsdldepplledgelfdvfvsvagvvgggaggagayagtayid-alaecr 1336  
QY 593 RAGG--VLAYELLPALDEVLASDSRFLLSWLEQA-----RAAAV 630  
Db 1337 ragglpvtsvawtpwig-tpaads---lgeqmsragltpldpaasldalaravgrragcv 1392  
QY 631 SEAEADFYEQNSRYQLT 647  
Db 1393 tvadidwerfasaytat 1409  
RESULT 7  
AAB30570  
ID AAB30570 standard; Protein; 871 AA.  
XX  
XX AAB30570;  
XX  
XX 19-MAR-2001 (first entry)  
XX  
XX A splice variant of a signal transduction polypeptide.  
DE  
XX Signal transduction; H19C5; kinase; cardiac disease; angina pectoris;  
KW congestive heart failure; dilated congestive cardiomyopathy;  
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
KW atherosclerosis; cardiac tumour; microbial infection; splice variant.  
XX



OS Homo sapiens.  
 PN WO200063381-A1.  
 PD 26-OCT-2000.  
 XX 11-APR-2000; 2000WO-US09488.  
 XX 16-APR-1999; 99US-0129553.  
 XX (SCIO-) SCIOS INC.  
 XX Zeng W, Stanton L, Kong H;  
 PI WPI; 2001-007013/01.  
 DR Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction -  
 XX  
 PS Claim 1; Page 74-76; 81pp; English.  
 XX  
 CC The present sequence represents a splice variant of human in signal  
 CC transduction polypeptide. The polypeptide is designated H19G5. The  
 CC protein is capable of regulating signal transduction and exhibits kinase  
 CC activity. The H19G5 transcript is expressed in the heart. H19G5  
 CC polypeptides and polynucleotides are useful for preventing or treating a  
 CC cardiac disease, such as congestive heart failure, dilated congestive  
 CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
 CC mitral valve disease, aortic valve disease or tricuspid valve disease,  
 CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
 CC arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis  
 CC and cardiac tumours in humans. The polypeptide is also useful for  
 CC detecting the expression of a protein capable of regulating signal  
 CC transduction or the expression of a phosphate group. The monoclonal antibodies can  
 CC be used as probes for detecting discrete antigens expressed by tissue or  
 CC cell samples, and therefore used in humans for localization and  
 CC monitoring of microbial infection.  
 XX  
 SQ Sequence 871 AA;

Query Match 3.0%; Score 118.5; DB 22; Length 871;  
 Best Local Similarity 18.6%; Pred. No. 0.24; Mismatches 307; Indels 319; Gaps 40;  
 Matches 165; Conservative 97;  
 QY 19 GGAAGDEAREAAVALVARL-LGPGPAAD---FSVSVERALAAKPGLD-----TY 65  
 DB 21 gsasqassqvsslrvgssqvgtepgpsldaegwtgeadlsdstptlqrpqeqatmrkf 80  
 QY 66 SLGG-GGAARVVRGCTGVAAGLHRYLRDFCG-----CHVAMSGSGLR----- 109  
 DB 81 slgrrgygavgygtfagdgag-----gmlgggpmwariawavsgseeeqearae 134  
 QY 110 -----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYSFVWMDWARWE 151  
 DB 135 sqseeqgeaearaeslpqsarvpvevgrptrssp-----eptpwe 175  
 QY 152 -----REI---DMMALNGINLAL-----AWSGQEAIWQRYLAL----- 182  
 DB 176 dlqgvslvgirdlsgdaaatisldisevdpaylnlsdydkylpfefmfrkvpksa 235  
 QY 183 -----GLTQAEINEF-----FTGP-----AFLANGRMGNLHTWDGP 213  
 DB 236 qpeppspmaeeelaefpeptwpvpgelgphaggleiteesedvdallaeaavgrkrkwssp 295  
 QY 214 LPPSWHIKQLYLRHVDQMSFGMTPLPAFAGHV-----PEATRVFPQVNVTKMG 266  
 DB 296 srsifhfg---rhlpdepaelqirrvkasvenhrlkgrpegkegpp---rrkpp 349  
 QY 267 SWGHFNCYSYCSFLAPEDPIFFIIGSLFLRELKIFGTGDIHYGADTFNEMQPPSSPS 326

DB 350 glasfrlsglkswdrap-----tfrel-----sdtv 378  
 QY 327 LAATTAATVVEAMTAVDTEAVWLLOGLWLOHQFQWGPQAQIRAVLCAVPRG--RLVLVLDLF 384  
 DB 379 lqgsvtlacqvsaqpaadgatskdaple-----sssrvlisatknfqltlv 430  
 QY 385 AESQPVYTRTASFOGQPFIMCMLHNFNGHGLFGALEAVNGGPEARLFNPSTVMVGTGMA 444  
 DB 431 aedlgvyt-----csvsnalgtvtgtvlr-----kaer--psss----- 463  
 QY 445 PEGISONEVWYSLMAELGWRKDPVPD-----LAAW--VTSPAARRYGVSHPDAGAA 493  
 DB 464 -----pcpdigevyadgvllywkpvesygpvtiyvqcsleggs 501  
 QY 494 WRLLRSVYNC-----SGEAC 509  
 DB 502 wtllasdlfdccyltsklrsgtytfrtactvskagmgypspseqvllggpshlasees 561  
 QY 510 RGHNRSPVLRPSLOMNTSIWYNR-SDVFEAWRLLLTSAPSLATSPAPRYDLDLTR--Q 566  
 DB 562 qgrsaqplpstktfafaqtqgrgrfsvrqcwekasgralaakiptyhpkdktavlreye 621  
 QY 567 AVQELVSLYEEARSAYLS-KELASLLRAGGVLAAYELLPALDEVLASDSRELLGSWLEQA 625  
 DB 622 alkgirhphlaqlhaaylsprhlvllel---csgpellplclae-rasyesevkdylwqm 678  
 QY 626 RAAA-----VSEAEADFVEQN---SRVOLTLMGPENILDYANKLAGLVANYTTPRWRL 677  
 DB 679 lsatqylhqnqhlhlrlsrnmiiteynll-----kvvdlnaqslsqekvlpdkfkf 732  
 QY 678 FLEALVDSVAOG---IPQOQHDKNVFOL-EQAFVLSKQRYPSQPRG 721  
 DB 733 yletnapellegqavp-----qtdlwaigvtafimlsaeypvsseg 774  
 RESULT 8  
 AAB30567  
 ID AAB30567 standard; Protein; 1351 AA.  
 XX  
 AC AAB30567;  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human signal transduction polypeptide.  
 XX  
 KW Signal transduction: H19G5; kinase; cardiac disease; angina pectoris;  
 KW congestive heart failure; dilated congestive cardiomyopathy;  
 KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
 KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
 KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
 KW atherosclerosis; cardiac tumour; microbial infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200063381-A1.  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 11-APR-2000; 2000WO-US09488.  
 XX  
 PR 16-APR-1999; 99US-0129553.  
 XX  
 XX (SCIO-) SCIOS INC.  
 PA  
 XX Zeng W, Stanton L, Kong H;  
 PI WPI; 2001-007013/01.  
 DR N-PSDB; AAC62285.  
 XX  
 PT Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction -  
 XX

PS Claim 1; Page 55-57; 81pp; English.

XX The present sequence represents a human protein with putative function

CC in signal transduction. The polypeptide is designated H19G5. The protein

CC is capable of regulating signal transduction and exhibits kinase

CC activity. The H19G5 transcript is expressed in the heart. H19G5

CC polypeptides and polynucleotides are useful for preventing or treating a

CC cardiac disease, such as congestive heart failure, dilated congestive

CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,

CC mitral valve disease, aortic valve disease or tricuspid valve disease,

CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,

CC arterial or rheumatic hypertension, arteriosclerosis, atherosclerosis

CC and cardiac tumours in humans. The polypeptide is also useful for

CC detecting the expression of a protein capable of regulating signal

CC or acceptor molecule of a phosphate group. The monoclonal antibodies can

CC be used as probes for detecting discrete antigens expressed by tissue or

CC cell samples, and therefore used in humans for localization and

CC monitoring of microbial infection.

XX

SQ Sequence 1351 AA;

Query Match 3.0%; Score 118.5; DB 22; Length 1351;

Best Local Similarity 18.6%; Pred. NO. 0.48;

Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAGDEAREAAVALVRL-LGCPAAD---FVSVERALAAKPGLD-----TY 65

DB 501 gasqssqvsirvgssqvtgpepsidaegwtgeaedlsdstptlqrpqeqatmrkf 560

QY 66 SLGG-GGAARVRVRSSTGVAAGALHRYLRDFCG-----CHVAGWSQLR----- 109

DB 561 slgrrgyagvgygtfagdgag-----gmlggpmwariawavsqseeeqeeareae 614

QY 110 -----LP-----RPLPAVPGELTEATPRYRYQNVCTQSYSFVWWDWARVE 151

DB 615 sqseeqgearaesplbqvaaarpvgraptrssp-----eptpwe 655

QY 152 -----RET---DWMALINGLAL-----AWSQEAIAWRYLAL----- 182

DB 656 digqvalqirdlsgdaaeadtsldisevdpaylnlsldydikylpfefmifkrkvpksa 715

QY 183 -----GLTQAEINEF-----FTGP-----AFLAWGRMGNLHTWDGP 213

DB 716 qppeppmaeeelaefpeptwppwpgelgphaglleiteesedvdallaeaavgrkrkwssp 775

QY 214 LPSPWHIKQLYLOHRVLDQMRSGMTPLVPAPAGHV-----PEAVTRVFPQVNVTKMG 266

DB 776 srslfhfpg---rhlpidepaeligirervkasvehisrlkgrpeglekgpp---rrkbp 829

QY 267 SWGHFNCSYSCSFLAPEDPIFIIGSLFLRELKKEFGTDHIYCADTFNEMQPPSPSEPSY 326

DB 830 glasfrlsglkswdrap-----tfirel-----sdetv 858

QY 327 LAATTAAYEAMTAVDTFAVWLQGMFLQHQPFQWGAQIYRAVLGAVPRG---RLVLVDLF 384

DB 859 lggsvtlacvsaqaapaqatkskdaple-----sssrviisatlnkfultilvv 910

QY 385 AESOPVYTRTASQGGQFICMCLHNFNGHGLFGALEAVNGGPEAARLPNPNMTVGTMGA 444

DB 911 aedlgvyt-----cvsvalgtvtttgvlr-----kaer---pss----- 943

QY 445 PEGISQNEVVYSLMAELGWRKDPVPD-----LAAM---VTSFAARRYGVSHPDAGAA 493

DB 944 -----pcpdlgevyadvllvwkpesygpvtivqcsleggs 981

QY 494 WRLLRSVYNC----- 1041

DB 982 wtllasdlfdccyltsklrsrgtgytftacvksagmgpyspseqlvgpshlasees 1041

QY 510 RGNRSPLVRRPSLQMTWTSWYNR-SDVFEAWRLLLTSAPSLATSPAFRYDLDDLTLR--Q 566

DB 1042 qgrsaqlpstktfqtqgrgrfsvrrcwekasgralaaakliipbkdktavireye 1101

QY 567 AVQELVSLYVEARSAYLS-KELASLLRAGGVLAYELLPALDEVLASDSRFLGSLWEQA 625

DB 1102 alkglrphlaqlhaaylsprhlvllel--csgpellpclae-rasyseevkdyiwqm 1158

QY 626 RAAA-----VSEAEADYEQN---SRVQLLWEGEINLQYANKQAGLVANYTPRWRL 677

DB 1159 lsatqylhnhqhlhldlrseumilteynll-----kvdldnagslsqekvlpsdkfk 1212

QY 678 FLEALVDSVAGQ---IPFQQHQFKNVFL-EOAFVLSKQRYSPQPRG 721

DB 1213 yletmapelleggavp-----qtdiwaigvtafimlseaypvsseg 1254

RESULT 9

AAB30568

ID AAB30568 standard; Protein; 1610 AA.

XX AAB30568;

XX 19-MAR-2001 (first entry)

XX A full length human signal transduction polypeptide.

DE Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;

XX congestive heart failure; dilated congestive cardiomyopathy;

KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;

KW mitral valve disease; aortic valve disease; tricuspid valve disease;

KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;

KW atherosclerosis; cardiac tumour; microbial infection.

XX Homo sapiens.

OS WO200063381-A1.

PN 26-OCT-2000.

PD 11-APR-2000; 2000WO-US09488.

PF 16-APR-1999; 99US-0129553.

PR (SCIO-) SCTOS INC.

PA Zeng W, Stanton L, Kong H;

PI WPI; 2001-007013/01.

DR N-PSDB; AAC62286.

XX Novel h19G5 polypeptides capable of regulating signal transduction and

PT exhibiting kinase activity useful for identifying antibodies to treat

PT cardiac diseases, and additional mediators of signal transduction

XX Claim 1; Page 61-65; 81pp; English.

XX The present sequence represents a human protein with putative function

CC in signal transduction. The polypeptide is designated H19G5. The protein

CC is capable of regulating signal transduction and exhibits kinase

CC activity. The H19G5 transcript is expressed in the heart. H19G5

CC polypeptides and polynucleotides are useful for preventing or treating a

CC cardiac disease, such as congestive heart failure, dilated congestive

CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,

CC mitral valve disease, aortic valve disease or tricuspid valve disease,

CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,

CC arterial or rheumatic hypertension, arteriosclerosis, atherosclerosis

CC and cardiac tumours in humans. The polypeptide is also useful for

CC detecting the expression of a protein capable of regulating signal

CC or acceptor molecule of a phosphate group. The monoclonal antibodies can

CC be used as probes for detecting discrete antigens expressed by tissue or

CC cell samples, and therefore used in humans for localization and

CC monitoring of microbial infection.

XX

SQ	Sequence	1610 AA;
XX	DT	25-SEP-2001 (first entry)
XX	DE	Human protein kinase SGK145.
XX	KW	Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human; antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;
XX	KW	analeptic; hypotensive; hypertensive; immunosuppressive; antiallergic;
XX	KW	antipsoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;
XX	KW	osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;
XX	KW	vasotropic; antidiabetic; gene therapy.
OS	OS	Homo sapiens.
XX	PN	WO200155356-A2.
XX	PD	02-AUG-2001.
XX	XX	25-JAN-2001; 2001WO-US02337.
XX	XX	25-JAN-2000; 2000US-0178078.
XX	XX	31-JAN-2000; 2000US-0179364.
XX	XX	17-FEB-2000; 2000US-0183173.
XX	XX	17-MAR-2000; 2000US-0190162.
XX	XX	29-MAR-2000; 2000US-0193404.
XX	XX	13-NOV-2000; 2000US-0247013.
XX	XX	(SUGE-) SUGEN INC.
XX	XX	Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;
XX	XX	WPI; 2001-476202/51.
XX	XX	N-PSDB; AAH46904.
XX	XX	Kinase polypeptides useful for treating cancers, Alzheimer's disease, viral infections, diabetes, obesity, organ transplant rejection and rheumatoid arthritis -
XX	XX	Claim 7; Page 215; 218pp; English.
XX	XX	The invention provides human protein kinases and protein kinase-like enzymes and polynucleotides encoding the polypeptides. The kinase polypeptides and their modulators are useful for treating a disease or disorder such as cancer, immune-related diseases, cardiovascular disease, brain or neuronal-associated disease and metabolic disorders, including cancers of tissues, cancers of hematopoietic origin, diseases of the central nervous system, diseases of the peripheral nervous system, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, viral infections, infections caused by prions, bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction, mood disorders, attention disorders, cognitive disorders, hypotension, metabolic disorders, and organ transplant rejection. They are also useful for treating rhinitis, autoimmunity, atherosclerosis, psoriasis, osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic inflammatory bowel disease, rheumatoid arthritis, metabolic disorders such as diabetes, obesity, cardiovascular diseases such as reperfusion injury, coronary thrombosis, clotting disorders and atherosclerosis, ocular diseases such as glaucoma, retinopathy and macular degeneration, psychiatric and neurological disorders such as anxiety, schizophrenia, dementia, manic depression, etc. The polynucleotides are useful in gene therapy techniques to treat the above mentioned disorders. Sequences AAB85491-85522 represent the human protein kinases of the invention.
XX	XX	Sequence 1618 AA;
XX	XX	Query Match 3.0%; Score 118.5; DB 22; Length 1618;
XX	XX	Best Local Similarity 18.6%; Pred. No. 0.63;
XX	XX	Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;
QY	19	GGAGDEAREAAVRAVLVRL-LGPGPAAD---FSVSVERALAAKPGLD-----TY 65
Db	760	gsasqassvslrgvssqteppslsdaegwtgeadlsdtpclqrpgqeatmrkf 819
QY	66	SLGG-GGAARVRVSGTGVAAAAGLHRYLRDFCG---CHVAMSGSQLR----- 109
Db	820	slgrrgyagvagytgfafggdag-----gmllggpmwariavavsqseeeqearae 873
QY	110	-----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSFVWWDVARWE 151
Db	874	sqseeqearaesplpqvsarpvpgvgraptrssp-----eptpwe 914
QY	152	-----REI--DWMLANGINLAL-----AMSGQEAIRWRYLAL----- 182
Db	915	dggvslvqlrldsgdaeaadtisldisevdpaylnlsldlydikylpfefmifkrvpksa 974
QY	183	-----GLTQAEINER-----FTGP-----AFLAWGMGNLHTWDGP 213
Db	975	qpeppspmaeeelaefpeptwppgelgphagleiteesedvdallaeaavgrkrkwssp 1034
QY	214	LPPSHIKOLYLOHRYLSDMRSEFGMPVLPAPAGHV-----PEAVTVFVQVNVTKMG 266
Db	1035	srsalfhfpq---rhlpdepaelgrervkasvehisrllkgrpegkegpp---rrkbp 1088
QY	267	SWGHFNCSYSCSFLAPEDIPFIIGSLFLRELIIKEFGTDHYGDTFNMOPPSSEPSY 326
Db	1089	glasrlslgkwdrap-----tflrel-----sdetvv 1117
QY	327	LAAATTAATVAMTAVTEAWVLLQWLFQHPQFWGPAQIRAVLGAVPRG---RLVLVLDLF 384
Db	1118	lqgsvtlacqvsapaqaatwkdgaple-----ssrvllisatlknflltilvvv 1169
QY	385	ASOPVYTTASFGQGFOTWCLHNFNGHGLFGALEAVNGGPEAARLPNSTVMVGTGMA 444
Db	1170	aedlgyvt-----csvsnalgtvtgtvlr-----kaer--psss----- 1202
QY	445	PEGISQNEVVYSILMELGWRKQDPVD-----LAAW--VTSPAARYGVSHPDAGAA 493
Db	1203	-----pcpdigevyadgvlvllvkpvesygpvtvlyvcslleggs 1240
QY	494	WRLLSRVYNC-----SGEAC 509
Db	1241	wtllasdifdcyltsklrsgtytfrtaacvskagmgysspsqvllgapshlasees 1300
QY	510	RGHNRSPLVRRPSLQNTSIWNR--SDVFEAWRLLLTSAPSLSATSPAFRYDLDLTLR--Q 566
Db	1301	qgrsaqplpstktfatqtlqgrfsvvrqwcakasgralaakiipypkdktaavlreye 1360
QY	567	AVQELVSLVYEARSAYLS-KELASILLRAGGVLAVELLPALDEVLASDRFLLSWLRQA 625
Db	1361	alkgrhphladhaaylsprhlvllel--csgpellipclae-rasyseevkdyllwqm 1417
QY	626	RAAA-----VSPAEDAFYEON---SRYQTLTWGPEGNLIDYANKLAGLVANYTTPRWRL 677
Db	1418	lsatylhnhqhlidlrseemlityeaynl-----kvvdignadslsqekvlpsdkfk 1471
QY	678	FLEALVDSVAQG---IPFQHQHFDKNVFQL-EQAFVLSKQRYPSQPRG 721
Db	1472	yietmapelledggavp-----qtdiwaigvtafimlsaeypvsseg 1513

RESULT 10

AAB85504

ID AAB85504 standard; protein; 1618 AA.

XX

AC AAB85504;

768	gasqassqvslrvsgsqvgtppgsldaeagwtqeaedlsdstptlqrpqeqatmrkf	827	KW	myocardial infarction; cardiac arrhythmia; arteriosclerosis; atherosclerosis; cardiac tumour; microbial infection; splice variant.		
66	SLGG-GGAARVRVGRSTGVAAAGLHRYLRDFCG-----CHVAMSGSQLR-----	109	XX	Homo sapiens.		
828	slggrggyagvagygtfagggdag-----gm1gggpmwariawavsqseeeqearae	881	OS	Key		
110	-----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYFVWMDWARWE	151	FH	Domain		
882	sqseeqgearaesplpqvsarpvpgvgrptrssp-----eptpwe	922	FT	Location/Qualifiers		
152	-----REI-----DWMALNGINLAL-----AWSQEAIMQWRYLAL-----	182	FT	325..504		
923	diggvslvirdlsgdaeadtisidisevdpaylnlsldlykylpftfemifkrkvpksa	982	FT	/note= "guanine nucleotide exchange factor domain"		
183	-----GLTQAEINEF-----FTGP-----AFLANGRMGNLHTWDGP	213	FT	1094..1351		
983	qpeppspmaeeelaefpeptwpggelgphaglaiteesedvdallaavaavgrkrkwssp	1042	FT	/note= "kinase domain"		
214	LPPSWHIKQLYLRVLDQMRSEGMTPLPAFAAGHV-----PEAVTRVFQVNVTKMG	266	PN	2301..2553		
1043	srslfhfpg---rhlpdpeaelgrrerkasvehisrllkgrpeglekgpp---rkkp	1096	XX	/note= "kinase domain"		
267	SWGHFNCYSCSFLAPEDPIPIIGSLFLRELKEFGTHIYGADTFNEMQPPSPSEPSY	326	XX	WO200063381-A1.		
1097	glasrlslgskwdrap-----tfirel-----sdtv	1125	XX	26-OCT-2000.		
327	LAATTAATVAVEMTAVTEAVWLLOGLFQHQPOFWGPAQIRAVLGAVPRG--RLVLVLDLF	384	XX	11-APR-2000; 2000MO-US09488.		
1126	lgdsvtlacvsqaqaatwskdgaple-----ssrvliisatlknflltilvv	1177	XX	16-APR-1999; 99US-0129553.		
385	AESQPVYTRTASFGQPPITWCMLEHNEGGNHLFGALEAVNGPEAARLPNSTMTVGTGMA	444	XX	(SCIO-) SCIOS INC.		
1178	aedlgvvt-----cvsvalgtvttgvlr-----kaer--psss-----	1210	XX	Zeng W, Stanton L, Kong H;		
445	PEGISQNEVYVSLMAELGWRKDPDP-----LAAW--VTSFAARRYGVSHPDAGAA	493	XX	WPI; 2001-007013/01.		
1211	-----pcpdigevyadgvlvwkpvesygvvtiyvqcsleggs	1248	XX	N-PSDB; AAC62287.		
494	WRLLLRSVYNC-----	509	XX	Novel h19G5 polypeptides capable of regulating signal transduction and exhibiting kinase activity useful for identifying antibodies to treat cardiac diseases, and additional mediators of signal transduction		
1249	wtllaedifdcyltsklrggtyfrtacyskagmpyspseqvllggpshlaseees	1308	XX	Claim 1; Page 68-74; 81pp; English.		
510	RGNRSPLEVRPSLQMTSITWYNR-SDVFEAWRLLTSLAPSLSATSAFYRDLDLDFR--Q	566	CC	The present sequence represents a splice variant of human in signal transduction polypeptide. The polypeptide is designated H19G5. The protein is capable of regulating signal transduction and exhibits kinase activity. The H19G5 transcript is expressed in the heart. H19G5 polypeptides and polynucleotides are useful for preventing or treating a cardiac disease, such as congestive heart failure, dilated congestive cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve disease, aortic valve disease or tricuspid valve disease, angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial or renovascular hypertension, arteriosclerosis, atherosclerosis and cardiac tumours in humans. The polypeptide is also useful for detecting the expression of a protein capable of regulating signal transduction or the expression of a protein capable of acting as a donor or acceptor molecule of a phosphate group. The monoclonal antibodies can be used as probes for detecting discrete antigens expressed by tissue or cell samples, and therefore used in humans for localization and monitoring of microbial infection.		
1309	qgrsaqlpstktfatqfqrfsrvrcwksagralaakiip/hpkdkktavireye	1368	CC	Sequence 2596 AA;		
567	AVQELSVLYEARSAYLS-KELASLLRAGGVLAYELLPALDEVLASDSRFLGLSWEQA	625	CC	Query Match 3.0%; Score 118.5; DB 22; Length 2596;		
1369	aikglrphlaqlhaayisprhlvllel--csgpellpclae-rasysevekdylwqm	1425	CC	Best Local Similarity 18.6%; Pred. No. 1.3;		
626	RAAA-----VSEAEADFEQN---SRYQLTLMGPEGNILDYANKQLAGLVANYTTPRWRL	677	CC	Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;		
1426	lsatqylhnhqhlhldirsenmitteynll-----kvvdlnaqsksqekvippsdkfd	1479	CC			
678	FLEALVDSVAQG---IPFQHQFDKNVFQL-EQAFVLSKORYPSQPRG	721	CC			
1480	yletmapelleggavp-----qtdiwaigvtafmlsaeypvsseg	1521	CC			
RESULT	11		QY	19	GGAAGDEAREAAARVALVARL-LGPGPAAD---FSVSVERALAAKPGLD-----TY	65
AAB30569			Db	1746	gsasqassqvsslrvgssqvgtepsidsaegwtqeaedlsdstptlqrpqeqatmrkf	1805
XX	AAB30569;		QY	66	SLGG-GGAARVRVGRSTGVAAAGLHRYLRDFCG-----CHVAMSGSQLR-----	109
AC			Db	1806	slggrggyagvagygtfagggdag-----gm1gggpmwariawavsqseeeqearae	1859
XX			QY	110	-----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYFVWMDWARWE	151
DT	19-MAR-2001 (first entry)		Db	1860	sqseeqgearaesplpqvsarpvpgvgrptrssp-----eptpwe	1900
XX			QY	152	-----REI-----DWMALNGINLAL-----AWSQEAIMQWRYLAL-----	182
DE	A splice variant of a signal transduction polypeptide.					
KW	Signal transduction; H19G5; kinase; cardiac disease; angina pectoris; congestive heart failure; dilated congestive cardiomyopathy; hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension; mitral valve disease; aortic valve disease; tricuspid valve disease;					



Db 351 ahmpttavihevqrf-----gdivpl 372

RESULT 13

AAR72377

ID AAR72377 standard; Protein; 497 AA.

XX AC

XX AAR72377;

XX DT 15-NOV-1995 (first entry)

XX XX

DE Human auxillary cytochrome P450 species 2D6 variant 2 protein.

XX KW Human cytochrome P450; amplification; PCR; primer; expression vector;

XX KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;

KW carcinogen; mutagen; liver metabolism.

XX XX

OS Homo sapiens.

XX XX

FH Key Location/Qualifiers

FT FT Misc-difference 296 /note= "Cys to Arg variation"

FT FT Misc-difference 486 /note= "Thr to Ser variation"

XX XX

PN EP644267-A.

XX XX

PD 22-MAR-1995.

XX XX

PF 20-JUL-1994; 94EP-0111298.

XX XX

PR 21-JUL-1993; 93JP-0180246.

PR 20-JUL-1993; 93JP-0201120.

PR 30-JUL-1993; 93JP-0208279.

XX XX

PA (HAYASHI) HAYASHI K.

PA (SUMO) SUMITOMO CHEM CO LTD.

XX XX

PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;

PI Yabusaki Y;

XX XX

DR WPI: 1995-116991/16.

XX N-PSDB; AAQ87731.

XX XX

PT Evaluation of safety of a chemical cpd. - using recombinant yeast

PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase

XX XX

PS Examples; Page 91-93; 124pp; English.

XX XX

CC The amino acid sequence of the human auxillary cytochrome P450 species

CC 2D6 variant 2. This variant contains variations at residues 296: Cys to

CC Arg and 486: Thr to Ser, caused by variations at bases 886: T to C and

CC 1457: C to G in the DNA sequence. The cDNA was amplified by PCR using

CC the primers AAQ87763-6. The product was cloned into the yeast

CC expression vectors pAAH5N or pAHRH to produce the vectors p2D6 variant 2

CC for the expression of the cytochrome P450 alone or p2D6R variant 2 for

CC co-expression with the yeast NADPH-P450 reductase.

CC The vectors are used in a method for evaluating the safety of a chemical

CC compound by reacting the chemical compound with recombinantly produced

CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),

CC 2E1 (AAQ87716), or 3A4 (AAQ87718), or their auxiliary species and

CC variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused

CC protein or in cell extracts, and analysing the resulting metabolite to

CC assess the safety of the chemical compound. The method is useful for

CC determining whether the chemical compound, or its metabolite, will be

CC converted into a carcinogenic or mutagenic form through metabolism in the

XX CC liver.

XX XX

Sequence 497 AA;



Db 351 ahmpyttavihevqrf-----gdivpl 372

Search completed: September 21, 2002, 12:26:27  
Job time: 162 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 12:23:45 : Search time 14.54 Seconds  
(without alignments)  
1248.158 Million cell updates/sec

Title: US-09-836-613-2  
Perfect score: 3939  
Sequence: 1 MEAVAVAAAVGVLLAGAGG.....VDLAKKIFLYKYPGWAGSW 743

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3939	100.0	743	4	US-09-077-354B-2
2	121	3.1	4472	2	US-08-804-227C-2
3	120	3.0	3729	2	US-08-804-227C-4
4	113	2.9	7257	3	US-09-335-409-5
5	113	2.9	7257	4	US-09-568-102-5
6	113	2.9	7257	4	US-09-567-969-5
7	113	2.9	7257	4	US-09-568-480-5
8	113	2.9	7257	4	US-09-568-486-5
9	113	2.9	7257	4	US-09-568-472-5
10	107.5	2.7	1621	4	US-08-972-927-3
11	107.5	2.7	1622	4	US-08-972-927-6
12	105	2.7	1203	4	US-09-075-272-4
13	105	2.7	4150	2	US-09-428-517-2
14	104.5	2.7	726	2	US-08-313-185-49
15	104.5	2.7	726	2	US-08-459-499-13
16	104.5	2.7	726	3	US-09-082-614A-49
17	104	2.6	581	3	US-09-221-235-8
18	104	2.6	581	3	US-09-221-928-8
19	104	2.6	581	4	US-09-221-527-8
20	104	2.6	581	4	US-09-221-236-8
21	104	2.6	581	4	US-09-221-416-8
22	104	2.6	581	4	US-09-221-245-8
23	104	2.6	581	4	US-09-163-115-8
24	104	2.6	581	4	US-09-221-528-8
25	104	2.6	581	4	US-09-593-553-8
26	104	2.6	581	4	US-09-221-237-8
27	104	2.6	677	4	US-08-836-567-2

28	104	2.6	1197	4	US-08-836-567-12
29	101	2.6	3519	4	US-09-428-517-4
30	101	2.6	5215	4	US-09-105-537-2
31	100.5	2.6	2628	4	US-09-413-814-11
32	100	2.5	1323	1	US-08-026-138E-4
33	100	2.5	1841	2	US-08-804-227C-6
34	99.5	2.5	719	1	US-07-943-843-4
35	99.5	2.5	719	1	US-08-347-003-4
36	99	2.5	729	2	US-08-313-185-50
37	99	2.5	729	2	US-08-459-499-14
38	99	2.5	729	3	US-09-082-614A-50
39	99	2.5	1230	2	US-08-968-542C-35
40	98.5	2.5	670	2	US-08-473-750-11
41	98.5	2.5	670	2	US-08-477-326-11
42	98	2.5	1611	2	US-08-804-227C-5
43	97	2.5	1336	2	US-08-231-193A-58
44	97	2.5	1336	2	US-08-486-273A-58
45	97	2.5	1336	3	US-08-940-086A-58

ALIGNMENTS

RESULT 1  
US-09-077-354B-2  
; Sequence 2, Application US/09077354B  
; Patent No. 6255096  
; GENERAL INFORMATION:  
; APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
; APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
; TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
; TITLE OF INVENTION: -N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SA  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
; STREET: 400 GARDEN CITY PLAZA  
; CITY: GARDEN CITY  
; STATE: NEW YORK  
; COUNTRY: UNITED STATES  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/09/077, 354B  
; APPLICATION NUMBER: US/09/077, 354B  
; FILING DATE: 22-APRIL-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/00747  
; FILING DATE: 22-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POKALSKY, ANN R.  
; REGISTRATION NUMBER: 34,697  
; REFERENCE/DOCKET NUMBER: 12416  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 516 742 4343  
; TELEFAX: 516 742 4366  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 743 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Potentially-glycosylated Asn site,  
; LOCATION: 261  
; FEATURE:  
; NAME/KEY: Potentially-glycosylated Asn site,  
; LOCATION: 272  
; FEATURE:  
; NAME/KEY: Potentially-glycosylated Asn site,  
; LOCATION: 435

FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 503  
LOCATION:  
FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 513  
LOCATION:  
FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 526  
LOCATION:  
FEATURE: Potentially-glycosylated Asn site,  
NAME/KEY: 532  
LOCATION:  
US-09-077-354B-2

Query Match 100.0%; Score 3939; DB 4; Length 743;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAVAVAAAGVLLLAGAGGAAGDEAREAAAVRALVRLGPGPAADFVSVERALAAKP 60  
DB 1 MEAVAVAAAGVLLLAGAGGAAGDEAREAAAVRALVRLGPGPAADFVSVERALAAKP 60  
QY 61 GLDTYSLGGGGAARVRVRSSTGYVAAAGLHRYLRDFCGCHVAMSGSQLRPLRPVAVGE 120  
DB 61 GLDTYSLGGGGAARVRVRSSTGYVAAAGLHRYLRDFCGCHVAMSGSQLRPLRPVAVGE 120  
QY 121 LPEATPNRYRYONVCTQSYFVWMDARWEREIDNMALINGINLAWSGOEAIWQVYL 180  
DB 121 LPEATPNRYRYONVCTQSYFVWMDARWEREIDNMALINGINLAWSGOEAIWQVYL 180  
QY 181 ALGLTOAEINEFTGPAFLAWGRGNLHTWDGPLPSPWHIKOLYLQHRVLDQMRSGMTP 240  
DB 181 ALGLTOAEINEFTGPAFLAWGRGNLHTWDGPLPSPWHIKOLYLQHRVLDQMRSGMTP 240  
QY 241 VLPFAAGHVPEAVTRVPQVNVTKMSGWHFNCSYSCSFLAPEDPIFFIIGSLFLRELI 300  
DB 241 VLPFAAGHVPEAVTRVPQVNVTKMSGWHFNCSYSCSFLAPEDPIFFIIGSLFLRELI 300  
QY 301 KEFTGDHIYGADTFNEMQPPSSPSYLAATAATVAYEAMTAVDTAEVNLQGWLFQHPQPF 360  
DB 301 KEFTGDHIYGADTFNEMQPPSSPSYLAATAATVAYEAMTAVDTAEVNLQGWLFQHPQPF 360  
QY 361 WGPQAIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGPFIWCMLNHFGNGHGLFAL 420  
DB 361 WGPQAIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFQGPFIWCMLNHFGNGHGLFAL 420  
QY 421 EAVNGGPEAARLPNSTWGTGMAPEGISQNEVYVSLMAELGWRKDPVPLAAWVTSFAA 480  
DB 421 EAVNGGPEAARLPNSTWGTGMAPEGISQNEVYVSLMAELGWRKDPVPLAAWVTSFAA 480  
QY 481 RRYGVSHPDAGAAWRLLLRSVYNGSGEACRGNRSPVLRPSLQMNSTSIWYNRSDVFEAW 540  
DB 481 RRYGVSHPDAGAAWRLLLRSVYNGSGEACRGNRSPVLRPSLQMNSTSIWYNRSDVFEAW 540  
QY 541 RLLTSPATSLATSPAFRYDLDLTRQAVQELVSLYYEARSAYLSKELASLLRAGGVLAY 600  
DB 541 RLLTSPATSLATSPAFRYDLDLTRQAVQELVSLYYEARSAYLSKELASLLRAGGVLAY 600  
QY 601 ELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFEQNSRYQLTLWGPENILYAN 660  
DB 601 ELLPALDEVLASDSRFLGSLWLEQARAAVSEAEADFEQNSRYQLTLWGPENILYAN 660  
QY 661 KQLAGLVANYTTPRWLFLEALVDSVAQIPFQHQDKNVFQLEQAFVLSKQRYPSQPR 720  
DB 661 KQLAGLVANYTTPRWLFLEALVDSVAQIPFQHQDKNVFQLEQAFVLSKQRYPSQPR 720  
QY 721 GDTVDLAKKIIFLYPGHAGSW 743  
DB 721 GDTVDLAKKIIFLYPGHAGSW 743

RESULT 2

US-08-804-227C-2  
; Sequence 2, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rosteck, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804.227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4472 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-2

Query Match 3.1%; Score 121; DB 2; Length 4472;  
Best Local Similarity 21.6%; Pred. No. 0.19;  
Matches 176; Conservative 71; Mismatches 280; Indels 288; Gaps 44;

QY 4 VAVAAV-----GVLLLAGAGGAAGDEAREAAVRAVRLG-----PGPAAD 47  
DB 1686 VDAVAVVGHSGSIAAATVAGALSLE--DAAAVVALRAGLIGRYLAGRGMAAVPLPAGE 1743  
QY 48 FSVSVERALAAKPGDLYSLGGGAA-----RVRVRSSTGYVAAAGLH----- 90  
DB 1744 ----VEAGLAKWPGVEVAANGPASTVVSDDRVAAGYVAVCOAGVQARLIPVDYASHS 1799  
QY 91 RYLRFECG-CHVAMSGSQLRPR-PLPAV-----PGLTEATPNRYRYONVCTQSYSFV 143  
DB 1800 RHVEDLKGELERVLRSIGIRPSRPVPCSTVAGEQGPGEV----- 1838  
QY 144 WWDARWER-----EIDWMLANGINL-----ALAWSGOEAIWQVYLA 181  
DB 1839 -FDAGYFNRNLNRNRFSAVVGGLLEBGRHRRFTEVSAHPVLVHAIQTAEEA--DRSVHA 1895  
QY 182 LGLTQAE---INEFFTGPAFLAWGRGNLHTWDGPLPSPWHIKOLYL-----QHRVLDOM 233  
DB 1896 TGTLRQDDSPHLLTSTA-EAWAHGATL-TWDPALPPG-HLTTLPTYPNHHYWLID-- 1950  
QY 234 RSEGMTVPLPAFAGHVPEAVTRVPQVNVTKMSGWHFNCSYSCSFLAPEDPIFF-IIG 292  
DB 1951 ----TTPTTATTTSQPTDAWR-----YRVT-----WK-----ALTSSVPRHSIG 1988  
QY 293 SLFLRELKKEFGTDHIYGADTFNEMQPPSSPSYLAATAATVAYEAMTAV-----D 342  
DB 1989 RCLL-----VAPPTDGLLGLTTLVLSERGASVARLEVPICARR 2028

QY 343 TEAVLQGLWFOHQPWGPQO-----IRAVLGAVPRGLLVLDLFAESQPVYTRTASQ 398  
Db 2029 AEVAELL-----KPSMESAGENTVVSLLGLVP-----STDVARTSTIALIQ 2070  
QY 399 -----GQPF-----INCMLE-----HNFNGNHLFGALEAVN-----GG-----PEAA 430  
Db 2071 AVSDIGVPAARWALTRRAVAVPGTPODAGLAQWFGFVAALELPDIWGLIDLPEA 2130  
QY 431 RLFPNSTMTGMAPEGISQ-----NEVYSLMAELGWRKDPVPLAAMWTSFAARRGVSHP 488  
Db 2131 ELTRTETSPOTPERLPQTNRRRALEAAVLAGRGEQDQAVRASGIYGRVSRAAA 2190  
QY 489 DAGAAWR-----LLRSVYNSGCEACRGNRSPLVRRPSLQMTSIWYNRSDVFEAWRLIL 544  
Db 2191 AGAASQWPGSTVLITGGMGAIGRR-----LARRLAE-----GAERLVL 2229  
QY 545 TS-----APSLATSPAFRDLID-----LTRAQVOELVSLV----- 575  
Db 2230 TSRGPEAPGA-----ELABELRGHCEYVHAACDVAERDALAALVTAYPPNAVFT 2282  
QY 576 -----YEARSAYLSKELASLLRAGGVLAYELLPALDEVLASDSRFL-----GSMLEQARA 627  
Db 2283 AGILDDAVDTLSPESFETVRGAKVCGAELLHQLTADIKGLDAFVFSVGTGWNAGOG 2342  
QY 628 A-AVSEAEADFEQNSR-----YQTLWGPEG 653  
Db 2343 AYAAAALDALAERRRAAGLPAATSVAMGLWGCGG 2377

RESULT 3  
US-08-804-227C-4  
; Sequence 4, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3729 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-4

Query Match 3.0%; Score 120; DB 2; Length 3729;  
Best Local Similarity 22.4%; Pred. No. 0.18;

Matches 192; Conservative 72; Mismatches 265; Indels 328; Gaps 47;  
QY 4 VAVAAV-----GVLLLAGAGGAGDEAREAAVRAVRLILG-----PGPAAD 47  
Db 668 VDVAAVGSHGSEGTAAATVAGALSLE-----DAAAVVALRAGLIGRYLAGRGMAAVALPAGE 725  
QY 48 FSVSVERALAAKPGLDITYSLGGGAA-----RVRVRGSGTVAAAGLH----- 90  
Db 726 -----VEAGLAKWPGVEVAANGPASTVVVSGDRRAVAVVACQEGVQARLIPVDYASHS 781  
QY 91 RYLDRFCG-CHVANGSGSOLRLPR-PLPAV-----PGEUTEATPNRYRYQNVCTQSYSFV 143  
Db 782 RHVEDLKGELERVLGIRSPRSPVPCVSTVAGEQGPV----- 820  
QY 144 WMDWARWER-----ETDWMALNGINL-----ALAWSGOEAIWQRYLA 181  
Db 821 -FDAGWFRNLNRNRFSAVVGGLLECHRRFIEVSAHPVLVHAIEQTAEAA-----DRSVHA 877  
QY 182 LGLTQAE-----INEFTGPAFLAWGMGNLHTWDGPLPPSWHIKOLYL-----QHRVLDMQ 233  
Db 878 TGLRRQDDSPHRLITSTA-EAWAHGATL-TWDPALPPG-HLTLPTYPFNHHHYWLD-- 932  
QY 234 RSFGMTPLVLPFA-----AGHYPEAVT--RVFPQVNVTKM--GSHGHFNCSYSCS 278  
Db 933 -----TTPTTATTQSTPDQNPADALPYKVSWRRLDQDLSLARLDGRW----- 978  
QY 279 FLAPEDPIPIIGSLFLRELK-----EFGTDH-----IYGADTFENEMQPPSS 322  
Db 979 LLVVPEASADPSVAEGVARELTARGATVESITVEPGADRSLRCLLVDAITERDEAGPLRG 1038  
QY 323 EPSYLA-----AATTAVYEAMTAVDTEA-VWLQGLWFOHQPWFG 362  
Db 1039 IVSLLLAGDHAGADGARVVPVAGLAASLALIQAAGDAGTEAGLW----- 1083  
QY 363 PAQIRAVLGAVPRGLLVLDLFAESQPV-----YTRTASFOGPPFTWCMHNFNGHGLFG- 418  
Db 1084 -AVTRGAVAAVPG-----DVPAPSOALLMGFGRVAGIE-LPHCWGGLDLPTGPGDSGF 1135  
QY 419 -----ALEAVNGGPEAARLFPNSTWVGT-GMAPEGISQNEVYVLSMAEL 461  
Db 1136 RQLAATLAGRPAEDQVALRA--SGAYGRRLVRAASAAGGADGWRPRG-----TVLVV 1184  
QY 462 GWRKDPVPDLAAWVTSFAARRY--GVSHP-----DAGAAWRLLRVSYNCSGACRGNHR 514  
Db 1185 GDTAEVAGPLVRWLLNGARRVTLGSLGSLPEELADVAARVTV-----APCDPADR 1236  
QY 515 SPLVRRPSLQMTSIWYNRSDVFEAWRLILTSAPSLATSPAFRYDLDLTROA----- 567  
Db 1237 PALRTLLAEQAPTAV-----LVAPPVPPTP-----LAEMTAEALATLSA 1277  
QY 568 -----VOELVSLYYE-----PARSAYLSKELASLL 592  
Db 1278 KGLVDRSLSLDEPDPFLLEDGELDAFVVSFVAGVWGGAGGAGTAAGTAYLD-ALAECR 1336  
QY 593 RAGG--VLAYELLPALDEVLASDSRFLGLSWLEQA-----RAAAV 630  
Db 1337 RAGGLPVTSAVTPWLG-TFAADS-----LGEQMSRAGITPLDPAASLDALARAVRRAGCV 1392  
QY 631 SEAEADFEQNSRYQLT 647  
Db 1393 TVADIDWERFASAYTAT 1409

RESULT 4  
US-09-335-409-5  
; Sequence 5, Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/335,409  
CURRENT FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-102-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKQLYLQHRVLDQMRSEFGMTVPVLPAPAGHVPEAVTRVFPQVNVTKM-GSWGHFN 272  
Db 4805 IPAEMPLQGVIIHAAGALDD-----GVLDEQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLLAPEDPIPIIGSLFLRELKKEFGTDHIYGADTFNEMOPSPSEPSYLAATT 332  
Db 4847 ---NLHETAGNDLAFVLFSSMSGLLSGAGSQSNYAAANTFLDALAAHRAEGLAAQSL 4902

QY 333 A-----VYEAMTAVDTAEAVWLLQGWLFQHPQFQWGPQAIQIRAVLG---AVPRGRL--L 379  
Db 4903 AWGPNWDGGMAGLSAA-----LQARLARHGMGALSQAQGTALLQALARPETQLGAM 4955

QY 380 VLDLFAESQPVYTRTASFOGPFIMCMLHNFNGHGLFGALEAVNGGPEAARLFPNSTMW 439  
Db 4956 SLDVRAASQ-----ASGAAPVPMVRALVRAEARHTAAGAAGAL-----AARL----- 4997

QY 440 GTGMAPEGISONEVYVSLMAE-----LQWR-----KDPVPLD-----AAWTSFAARR 482  
Db 4998 --GALPEARRADEVKVVQAEIARVLSWSAASAVPVDRLSDGLDLSLTAVELRVNGOR 5055

QY 483 YGVSS-----HPDAGAAWRLLLRSVYNCSEACRGHNRSPV-----RRP 521  
Db 5056 VGATLPATLAFDHTVDALTRWLLDKVLAEPVSSAKSSPQVALDEPIAIGICRFP 5115

QY 522 SLOMNTSIWNRSDVEAEWRLLTSAPSLATSPAFRYDLDLTLRQAVQLVSLYE---E 578  
Db 5116 G-----GVADPEFNRLLLEEGSDAVVEPHERWD-----IDAFYDPPDP 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSWLEQARAA 628  
Db 5155 VRGKMTI-----RFGGFLSDIDRDPAPFGISPREATTMDPQORLLLETSWEAPEFAG 5207

QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 6  
US-09-567-969-5  
Sequence 5, Application US/09567969  
Patent No. 6355457  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,969  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-567-969-5

APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/335,409  
CURRENT FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-5

Query Match 2.9%; Score 113; DB 3; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKQLYLQHRVLDQMRSEFGMTVPVLPAPAGHVPEAVTRVFPQVNVTKM-GSWGHFN 272  
Db 4805 IPAEMPLQGVIIHAAGALDD-----GVLDEQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLLAPEDPIPIIGSLFLRELKKEFGTDHIYGADTFNEMOPSPSEPSYLAATT 332  
Db 4847 ---NLHETAGNDLAFVLFSSMSGLLSGAGSQSNYAAANTFLDALAAHRAEGLAAQSL 4902

QY 333 A-----VYEAMTAVDTAEAVWLLQGWLFQHPQFQWGPQAIQIRAVLG---AVPRGRL--L 379  
Db 4903 AWGPNWDGGMAGLSAA-----LQARLARHGMGALSQAQGTALLQALARPETQLGAM 4955

QY 380 VLDLFAESQPVYTRTASFOGPFIMCMLHNFNGHGLFGALEAVNGGPEAARLFPNSTMW 439  
Db 4956 SLDVRAASQ-----ASGAAPVPMVRALVRAEARHTAAGAAGAL-----AARL----- 4997

QY 440 GTGMAPEGISONEVYVSLMAE-----LQWR-----KDPVPLD-----AAWTSFAARR 482  
Db 4998 --GALPEARRADEVKVVQAEIARVLSWSAASAVPVDRLSDGLDLSLTAVELRVNGOR 5055

QY 483 YGVSS-----HPDAGAAWRLLLRSVYNCSEACRGHNRSPV-----RRP 521  
Db 5056 VGATLPATLAFDHTVDALTRWLLDKVLAEPVSSAKSSPQVALDEPIAIGICRFP 5115

QY 522 SLOMNTSIWNRSDVEAEWRLLTSAPSLATSPAFRYDLDLTLRQAVQLVSLYE---E 578  
Db 5116 G-----GVADPEFNRLLLEEGSDAVVEPHERWD-----IDAFYDPPDP 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSWLEQARAA 628  
Db 5155 VRGKMTI-----RFGGFLSDIDRDPAPFGISPREATTMDPQORLLLETSWEAPEFAG 5207

QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 5  
US-09-568-102-5  
Sequence 5, Application US/09568102  
Patent No. 6346404  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,102  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6; Mismatches 183; Indels 142; Gaps 22;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKOLYQHRVLDQMRSGMTPVLPAPAGHVPEAVTRFVQVNVTKM-GSWGHN 272  
Db 4805 IFAEMPLOGVHAAGALDD-----GVLDQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLAPEDPFIPIIGSLFLRELKEFGTDHIYGADTFNEMOPSPSEPSYLAAT 332  
Db 4847 ---NLHETAGNDLFAFFVLFS-SMSGLLGSAGOSNYAAANTFLDALAAHRAEGLAAQSL 4902

QY 333 A-----VYEMTAVDTEAVWLQGWLFQHOPOFGWPAQIRAVLG---AVPRGL--L 379  
Db 4903 ANGPNSDGCGMAAGLSAA-----LQARLARHGMGALSQAQGTALLGQALARPETQLGAM 4955

QY 380 VLDLFAESOPVYTRTASFOGQPFIMCMLHNFNGHGLFGALEAVNGGPEAARLFPNSTMV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPMRALVRAEARTHAAGAOGAL-----AARL----- 4997

QY 440 GTGMAPEGISQNEVYSLMAE---LGWR-----KDPVPL-----AAWVTSFAARR 482  
Db 4998 --GALPEARRADEVRKVQAEIARVLSWSAASAVPVDRLSDGLDLSLTAVELRNVLGQR 5055

QY 483 YGVS-----HPDAGAAWRLLLRSVYNCSEACRGNRSLV-----RRP 521  
Db 5056 VGATLPATLAFDHPVDALTRWLLDKVLAVAPSVSSAKSPQVALDEPIAIGICRPP 5115

QY 522 SLOMNTSIWYNSRSDVFEAWRLLTSAPSLSATSPAFRYDLDLTLROAVQELVSLYE--E 578  
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPDPD 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSWLEQARAA 628  
Db 5155 VRGKMT-----RFGGLSDIDRDPFAFGISPREATTMDPQORLLLETSSWEAFERAG 5207

QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 7  
US-09-568-480-5  
; Sequence 5, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,480  
; PRIOR FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 7257  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-480-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6; Mismatches 183; Indels 142; Gaps 22;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKOLYQHRVLDQMRSGMTPVLPAPAGHVPEAVTRFVQVNVTKM-GSWGHN 272  
Db 4805 IFAEMPLOGVHAAGALDD-----GVLDQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLAPEDPFIPIIGSLFLRELKEFGTDHIYGADTFNEMOPSPSEPSYLAAT 332  
Db 4847 ---NLHETAGNDLFAFFVLFS-SMSGLLGSAGOSNYAAANTFLDALAAHRAEGLAAQSL 4902

QY 333 A-----VYEMTAVDTEAVWLQGWLFQHOPOFGWPAQIRAVLG---AVPRGL--L 379  
Db 4903 ANGPNSDGCGMAAGLSAA-----LQARLARHGMGALSQAQGTALLGQALARPETQLGAM 4955

QY 380 VLDLFAESOPVYTRTASFOGQPFIMCMLHNFNGHGLFGALEAVNGGPEAARLFPNSTMV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPMRALVRAEARTHAAGAOGAL-----AARL----- 4997

QY 440 GTGMAPEGISQNEVYSLMAE---LGWR-----KDPVPL-----AAWVTSFAARR 482  
Db 4998 --GALPEARRADEVRKVQAEIARVLSWSAASAVPVDRLSDGLDLSLTAVELRNVLGQR 5055

QY 483 YGVS-----HPDAGAAWRLLLRSVYNCSEACRGNRSLV-----RRP 521  
Db 5056 VGATLPATLAFDHPVDALTRWLLDKVLAVAPSVSSAKSPQVALDEPIAIGICRPP 5115

QY 522 SLOMNTSIWYNSRSDVFEAWRLLTSAPSLSATSPAFRYDLDLTLROAVQELVSLYE--E 578  
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPDPD 5154

QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSWLEQARAA 628  
Db 5155 VRGKMT-----RFGGLSDIDRDPFAFGISPREATTMDPQORLLLETSSWEAFERAG 5207

QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 8  
US-09-568-486-5  
; Sequence 5, Application US/09568486  
; Patent No. 6355459  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,486  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 7257  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-486-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2.6; Mismatches 183; Indels 142; Gaps 22;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSWHIKOLYQHRVLDQMRSGMTPVLPAPAGHVPEAVTRFVQVNVTKM-GSWGHN 272  
Db 4805 IFAEMPLOGVHAAGALDD-----GVLDQTTDRFSRVLPKVTGAW----- 4846

QY 273 CSYSCSFLAPEDPFIPIIGSLFLRELKEFGTDHIYGADTFNEMOPSPSEPSYLAAT 332  
Db 4847 ---NLHETAGNDLFAFFVLFS-SMSGLLGSAGOSNYAAANTFLDALAAHRAEGLAAQSL 4902

Db 4847 ---NLHELTAAGNDLAFVFLFS-SMSGLLGSAGQSNYAANTFLDALAAHRAEGLAAQSL 4902  
QY 333 A-----VYEAMTAVDTAEVWLLQGWLFQHQPOFGWPAQIRAVLG---AVPRGRL--L 379  
Db 4903 AWGPNWDGMAAGLSAA-----LQARLARHGMGALSQAQGTALLGQALARPETQLGAM 4955  
QY 380 VLDLFAESQPVYTRTASFQGOFFIWMCLHNFNGHGLFGALEAVNGGPEAARLFPNSTMV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPPVNRALVRAEARHTAAGAOGAL-----AARL----- 4997  
QY 440 GTGMAPEGISQNEVVYSLMAE-----LGWR-----KDPVPDL-----AAWVTSAARR 482  
Db 4998 --GALPEARRADEVRKVVQAEIARVLSMSAASAVPVDRLPLSLDGLDLSLTAVELRNVLGQR 5055  
QY 483 YGVY-----HPDAGAAWRLLLRSVYNCSEACRGHNRSLV-----RRP 521  
Db 5056 VGATLPATLAFPHPTVDALTRWLLDKVLAVAEPSVSSAKSSPQVALDEPIAIIIGICREP 5115  
QY 522 SLOMNTSIWYNSDYFEAWRLLLTSAPSLSATSPAFRYDLDLTLTRQAVQELVSLYE---E 578  
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPPDP 5154  
QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSLWLEQARAA 628  
Db 5155 VRGKWT-----RFGGFLSDIDRDPAPFGISPREATTMDPQORLLLETSSWEAFERAG 5207  
QY 629 AVSE 632  
Db 5208 ILPE 5211  
RESULT 9  
US-09-568-472-5  
; Sequence 5, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 7257  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-5  
Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 2,6;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;  
QY 214 LPPSWHIKOLYLOHRVLDQMRSGMTFVLPAFAGHVPEAVTRVFPQVNTKMG-SWGHFN 272  
Db 4805 IPAENPLQGVIIHAAGALDD-----GVLDQETTDFRSRLVAPKVTGAW---- 4846  
QY 273 CSVSCSFLAPEDIPPIITGLFLRELKEFGDHIYAGDTFNMQPPSPSEPSYLAANA 332  
Db 4847 ---NLHELTAAGNDLAFVFLFS-SMSGLLGSAGQSNYAANTFLDALAAHRAEGLAAQSL 4902  
QY 333 A-----VYEAMTAVDTAEVWLLQGWLFQHQPOFGWPAQIRAVLG---AVPRGRL--L 379  
Db 4903 AWGPNWDGMAAGLSAA-----LQARLARHGMGALSQAQGTALLGQALARPETQLGAM 4955

QY 380 VLDLFAESQPVYTRTASFQGOFFIWMCLHNFNGHGLFGALEAVNGGPEAARLFPNSTMV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPPVNRALVRAEARHTAAGAOGAL-----AARL----- 4997  
QY 440 GTGMAPEGISQNEVVYSLMAE-----LGWR-----KDPVPDL-----AAWVTSAARR 482  
Db 4998 --GALPEARRADEVRKVVQAEIARVLSMSAASAVPVDRLPLSLDGLDLSLTAVELRNVLGQR 5055  
QY 483 YGVY-----HPDAGAAWRLLLRSVYNCSEACRGHNRSLV-----RRP 521  
Db 5056 VGATLPATLAFPHPTVDALTRWLLDKVLAVAEPSVSSAKSSPQVALDEPIAIIIGICREP 5115  
QY 522 SLOMNTSIWYNSDYFEAWRLLLTSAPSLSATSPAFRYDLDLTLTRQAVQELVSLYE---E 578  
Db 5116 G-----GVADPESEFWRLLEEGSDAVVEVPHERWD-----IDAFYDPPDP 5154  
QY 579 ARSAYLSKELASLLRAGGVLA-----YELLPALDEVLASDSRFL--GSLWLEQARAA 628  
Db 5155 VRGKWT-----RFGGFLSDIDRDPAPFGISPREATTMDPQORLLLETSSWEAFERAG 5207  
QY 629 AVSE 632  
Db 5208 ILPE 5211  
RESULT 10  
US-08-972-927-3  
; Sequence 3, Application US/08972927  
; Patent No. 6166290  
; GENERAL INFORMATION:  
; APPLICANT: Rea, Phillip A  
; APPLICANT: Lu, Yu-Ping  
; APPLICANT: Li, Ze-Sheng  
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN  
; TITLE OF INVENTION: PLANTS  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C., 22nd  
; STREET: One Commerce Square, 2005 Market Street, 22nd  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: US  
; ZIP: 19103-7086  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/972,927  
; FILING DATE: 18-NOV-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/031,040  
; FILING DATE: 18-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/061,328  
; FILING DATE: 08-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Doyle Leary Ph.D., Kathryn  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: 9596-1202  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-965-1284  
; TELEFAX: 215-567-2991  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1621 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-972-927-3

Query Match 2.7%; Score 107.5; DB 4; Length 1621;  
Best Local Similarity 19.8%; Pred. No. 0.72;  
Matches 126; Conservative 75; Mismatches 214; Indels 221; Gaps 33;

QY 175 WQVYLALG---LQAEINEFTGPAFLAWGRMGNL-----HTWDGPL--PPSHHIKOL 223  
DB 241 WNPMLTGLSKRPLTEKQVYLDT-----WDQETLFTSFQHSWDKELQKPPMLLRAL 294  
QY 224 YLQHRVLDQMRSGMTPLVPAFAGHVPFAVTRVFPQVNVTKMGs---WGHF--NCYSVCSF 279  
DB 295 N-----NSLGGRFWGGFWKIGNDCS- 315  
QY 280 LLAPEDPIPIIGSLFLRLIIEFGTD-----HIYGADTFNEMQPPSSPSYLA 328  
DB 316 -----QFVGPLLLNQLKSMQEDAPAMGYIYAFSIFGVVFGVL---CEAQYFQ 362  
QY 329 AATTAVYEAMTAVDEAVWLLQGWLFQHQPFQWGPQAIIRAVLGAVPRGRLVLDLFAES- 387  
DB 363 NVWRVGYRLRSA-----LIAAVERKSLRTNERRKFEQTKIT--NLMTD--AESL 410  
QY 388 QPVYTRTASFQOPF-----IWCMLHFNFGNHLFGALEAVNGGPEAARLFPNSTMVGTGM 443  
DB 411 QOICOSLHTMWSAPFRIIIALILLYQQLGVASLIGALLV-----LMFPLQTVLIISKM 463  
QY 444 ---APEGISONEVVVSLMAELGWRKDPVDPDLAAAVTSPARRYGVSHPDAGAAWRLLLRS 500  
DB 464 QKLTEGQRTDKRIGLMEVLAAMDTV--KCYAWENSE----- 500  
QY 501 VYNCSEACRGNHNSPLVRPRLSQMNTSIWYNSRSDVFEAWRL--LTSAPSLATSPAFR-Y 558  
DB 501 -----QSKVQTVRDELS-----WFKSOLLGALNMFILNSIPVLTIVISFGVF 544  
QY 559 DLL--DLTRQAVOELVSLYIEARSAYLSKELASLLR-----AGGVLAYELLPAL 606  
DB 545 TLGGDLT-----PAR-AFTSLSLFVAVLRFPLEMLPNIITOVNANVSLKRL 590  
QY 607 DEVLASDSRFLGSLWLEQARAAVSEAEADFY-----EQNSRYQLTLCGPEGNILDYANK 561  
DB 591 EEVLAETEERILLPNPIEPGEPAISIRNGYFSWDSKGDRTLSNINLNDVPLGSLVAVVGS 650  
QY 662 -----QLAGLVANYTPR--WRLFLALVDSVAQGIPIFOQH 695  
DB 651 TGEKTSLSIALGELPATSDAIVTLRGSVA--YVPQVSW-IFNATVRDNIILFGSPFDEE 707  
QY 696 QFDK--NVFQLEQAVLSKQRYPSOPRGDTVDLAKK 729  
DB 708 KYERAIDVTSKHDLEL-----LPGGDLTEIGER 736

RESULT 11  
US-08-972-927-6  
Sequence 6, Application US/08972927  
Patent No. 6166290  
GENERAL INFORMATION:  
APPLICANT: Rea, Philip A  
APPLICANT: Lu, Yu-Ping  
APPLICANT: Li, Ze-Sheng  
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN  
NUMBER OF INVENTION: PLANTS  
CORRESPONDENCE ADDRESSES:  
STREET: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
STREET: Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: US  
ZIP: 19103-7086  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,927  
FILING DATE: 18-NOV-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,040  
FILING DATE: 18-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/061,328  
FILING DATE: 08-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Doyle Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: 9596-12U2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-965-1284  
TELEFAX: 215-567-2991  
TELEX: 831-494  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1622 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-972-927-6

Query Match 2.7%; Score 107.5; DB 4; Length 1622;  
Best Local Similarity 21.1%; Pred. No. 0.72;  
Matches 116; Conservative 63; Mismatches 167; Indels 203; Gaps 30;  
QY 268 WGHF--NCYSVCSFLLAPEDPIPIIGSLFLRLIIEFGTDHIYADTFNEMQPPSSPSY 326  
DB 304 WGGFWKIGNDCS-----QFVGPLLLNELLK-----SMQ--LINEPAW 337  
QY 327 LAAATVAYEAMTAVDTEAVWLLQGWLFQHQPFQWGPQAIIRAVLGAVPR-----GR- 377  
DB 338 IG-----YIYAFSIFGVVGLGVLCSEAQYFQNVNRV--GYRLRSALIAAVERKSLRTNEGRK 392  
QY 378 -----LLVLDLFAES--QPVYTRTASFQOPF-----IWCMLHFNFGNHLFGALEAV 423  
DB 393 KFOQTKIINLMTD--AESLQOICOSLHTMWSAPRIIVLVLLYQQLGVASIIIGALFLV 450  
QY 424 NGGPEAARLFPNSTMV---GTGMAPEGISQNEVVVSLMAELGWRKDPVDPDLAAAVTSPFAA 480  
DB 451 -----LMFPQTQTVIISKTKQKTKEGLQRTDKRIGLNEVLAAMDTV--KCYAWENSEF-- 500  
QY 481 RRYGVSHPDAGAAWRLLLRSVYNCSEACRGNHNSPLVRPRLSQMNTSIWYNSRSDVFEAW 540  
DB 501 -----QSKVQTVRDELS-----WFKKAQLLSAF 524  
QY 541 RL-LTSAPSLATSPAFR-YDLL--DLTRQAVOELVSLYIEARSAYLSKELASLLR--- 593  
DB 525 NMFILNSIPVLTIVISFGVSLGGDLT-----PAR-AFTSLSLFVSIVRPL 570  
QY 594 -----AGGVLAYELLPALDEVLASDSRFLG-----SWLQARAA 628  
DB 571 FMLPNIITQMVNANVSLNRLEVLSTEERVLPPNPPIEGOPAIIRNGYFSWDSKADRP 630  
QY 629 AVSEAEADFYEQNSRYQLTLCGPEGNIL-----DYANKOLA 664  
DB 631 TLSNINLDI-----PLGSLVAVVSGTGEKTSLSIALGELPARSDATVTLR 677  
QY 665 GLVANYTPR--WRLFLALVDSVAQGIPIFOQHOFDK--NVFQLEQAVLSKQRYPSOPR 720  
DB 678 GSVA--YVPQVSW-IFNATVRDNIILFGAPFDOEKYERVIDTALQHDLEL-----LPG 727  
QY 721 GDTVDLAKK 729





Db 3457 DAELHTVLPALSS-----WRRRRVHRRLODRYRVE---WKPEPAALDEVLGGWLFVVP 3510  
QY 375 RG-----RLLVLDLFAESQPVYTRTASFOQPFIMWMLHNFNGN 413  
Db 3511 RGLADGVVAVVAATARGGEVSVVELDPTDRDRAYAEAVAGRGVGVSVFSLWDDRR 3570  
QY 414 HGLFGALEAVNGGPEARLPNSTVMTGTMAPEG-----ISONEVVYSLMAELGWRKDPV 468  
Db 3571 HSEHSVVA-----GLAASLVLAQALVDLGRVGGPRLWLVTGCAVAG-PSDAGVVIDPV 3625  
QY 469 PDLAANVTFAARRYGVSHPDAGAAWRLLRSVYNGSGEACR-----GHNRSPLVR 519  
Db 3626 ---QAQWGF-GRVLGLEHPE---LWGLVDLPVGVGDEEVCRRFVGWASAGFEDQAVR 3678  
QY 520 RPSLOMNTSIWYNR 533  
Db 3679 -----GSGVWVR 3686

RESULT 14  
US-08-313-185-49  
; Sequence 49, Application US/08313185  
; Patent No. 5851763  
; GENERAL INFORMATION:  
; APPLICANT: Heym, Beate  
; APPLICANT: Cole, Stewart  
; APPLICANT: Young, Douglas  
; APPLICANT: Zhang, Ying  
; APPLICANT: Honore, Nadine  
; APPLICANT: Telenti, Amalio  
; APPLICANT: Bodmer, Thomas  
; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
; TITLE OF INVENTION: in Mycobacterium Tuberculosis  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/313.185  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 02356.0068-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 726 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-313-185-49

Query Match 2.7%; Score 104.5; DB 2; Length 726;  
Best Local Similarity 18.3%; Pred. No. 0.36;  
Matches 125; Conservative 66; Mismatches 189; Indels 303; Gaps 32;  
QY 86 AAGLHRYLDFCGCHVANGSQLRLPRPLPAVPGELTEATNR-----YR----- 130

Db 26 SAGAGTTTRD-----WPNQLRV-----DLLNQHNSNRNPLGDFDYRKFEFSL 69  
QY 131 -YY-----QNYCTQSYSFVWMDWARWEREIDWMALNGINLALAWS----- 169  
Db 70 DYYGLKDKALLTESQP--WW-----PADWGSVAGLFIRMAWHGAGTYSIDRGGA 120  
QY 170 -----GQEAIWQRYVLAGLTOAEINEFFFTGP 196  
Db 121 GRGOQREAPLNSWPDNVSIDKARRLLWPIKQYKOKISWADLFILAGNVALENSGFT-- 178  
QY 197 AFLANGMGNLHTDGPPLPPSWHIKOLYLOHRLVLDQMRSGMTPLVPAPAGHVPEAVTRV 256  
Db 179 --FGG-AGREDVWEPDLVDNMGDEKALUTHR-----H-PEALAKA 215  
QY 257 FPOVNVTKMGSMGFHCYSYCSFLLAPEDPIPIIIGSLFLRELKFEFGTDHYGADTFNE 316  
Db 216 --PLGATEMG-----LIYVNEGP-----DH----- 234  
QY 317 MQPPSEPSYLAATYAYEAMTAVDTAVWLLQGLWLFQHQPFQWGPQAIQIRAVLGAVPRG 376  
Db 235 ----SGEPLSAAAIRATFGNMGMDDETVALIAGHTLGKTHGAGPT---SNVGPDP-- 285  
QY 377 RLLVLDLFAESQPVYTR-----TASFGQGFPIWCMHLNFGNH---GLF-- 417  
Db 286 -----EAAPIEEQGLHWNASTYSGVGADAITSGLEVWVTQTPTOWNSITFFENLFKY 336  
QY 418 -----GAL--EAVNGGPEARLPNSTVMTGTMAPEGISQNEVVYSLMAELGWRKD 466  
Db 337 EWWQTRSPAGAIQFEAV-----DAPEIIPD-----PFDPSSKKRKPMTMLVTDLTFRD 383  
QY 467 P-----VPDLAAWVTSFAARYGVSHPDAGAAWRLLLSRVYNGSGEACRGNRSPL 517  
Db 384 PEFEXISRFLNDPOAFNEAFARAWFKLTHRDMGPKSYI----- 423  
QY 518 VRPSLOMNTSIWYNRSDVFEAWRLLLTASPSLATSPAFRYDLDL-----TROAVOEL 571  
Db 424 --GPEVPKEDLIQD-----PLQPIYNPT-EQDIIDLKFAIADSGLSVSEL 467  
QY 572 VSLYYEARSAYLSKELASLLRAG-----GVLAYELLPALDEVL 610  
Db 468 VSV-----AWAS---ASTFRGDKRGANGARLALMPQRDWDVNAAAVRAALPVLEKIQ 517  
QY 611 ASDSRFLGSLWLEQARAAAASEA 633  
Db 518 KESGKASLADIIVLAGVVGVEKA 540

RESULT 15  
US-08-459-499-13  
; Sequence 13, Application US/08459499  
; Patent No. 5871912  
; GENERAL INFORMATION:  
; APPLICANT: Heym, Beate  
; APPLICANT: Cole, Stewart T.  
; APPLICANT: Young, Douglas B.  
; APPLICANT: Zhang, Ying  
; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoni  
; TITLE OF INVENTION: Amended)  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.3  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,499  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/875,940  
FILING DATE: 30-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/929,206  
FILING DATE: 27-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/029,655  
FILING DATE: 11-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0110-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-459-499-13

Query Match 2.7%; Score 104.5; DB 2; Length 726;  
Best Local Similarity 18.3%; Pred. No. 0.36;  
Matches 125; Conservative 66; Mismatches 189; Indels 303; Gaps 32;

QY 86 \_AGLHRYLRDFCCCHVANGSQLRPLRPVAPVGETEATPNR-----YR----- 130  
DB 26 SAGAGTTTRD-----WPNQLRV-----DLLNQHSNRNPLGEDFDYRKEFSKL 69  
QY 131 -YY-----QNVCTQSYFVWMDWARWEREIDWMLNGINLALAWS----- 169  
DB 70 DYGLKDKLKALTESQP--WW-----PADWGSYAGLFIRMAWHGAGTYRSIDGRGA 120  
QY 170 -----GQBAIWQRYVYALGLTQAEINEFFTG 196  
DB 121 GRGQORFAPLNPVNDVKARLLWPIKQYQKISWADLFILAGNVALENSGERT-- 178  
QY 197 AFLAWGRMGNLHTWDGFLPPSHIKQLYLQHRVLDQMRSEGTTPVLPAPAGHVPEAVTRV 256  
DB 179 --FGFG-AGREDVWEPLDYNWGDERAWLTHR-----H--PEALAKA 215  
QY 257 FPOVNTKMGSMGHFNCSYSCSFLAPEDPFIPIIGSLFLRELKEFGTDHIYGADTFNE 316  
DB 216 --PLGATEMG-----LIYVNEGP-----DH----- 234  
QY 317 MQPPSEPSVLAATAYEAMTAVTEAVMLQGNLFQHQPOFGWPAQIRAVLGAVPRG 376  
DB 235 ----SGEPLSAAAIATFGNMGNDIEETVALIAGGHTLGKTHGAGPT---SNVGGDP-- 285  
QY 377 RLVLVDLFAESQPVYTR-----TASFOGQPFICWMLHNFNGNH---GLF-- 417  
DB 286 -----EAPFIEQGLWASTYSGVGCADAITSGLEVVTWTQTQWSNYFFENLFKY 336  
QY 418 -----GAL--EAVNGGPEAARLPNSTVMVTGMAPEGISONEVYVYSLMAELGWRKD 466  
DB 337 EMQVTRSPACAIQFEAV---DAPEIIPD-----PDPSPKRRKPTMLVTDLTIRFD 383  
QY 467 P-----VPDLAAVWTSFAARRYGVSHPDAGAAWRLLLRSYNCSGEACRGNRSPL 517  
DB 384 PEFKISRRFLNDPOAFNEAFARAWFKLTHRDMGPKSRYI----- 423  
QY 518 VRRPSLQMNSTSWNRSRDFEAWRLLLTSAPSLATSPAFRYDLIDL-----TRQAVQEL 571  
DB 424 --GPEVPEKDLINQD-----PLPOPIYNPT--EQDIIDLKFAIDAGLSVSEL 467

QY 572 VSLYYEEARSAYLSKELASLLRAG-----GVLAYELLPALDEVL 610  
DB 468 VSV-----AWAS-----ASTFRGGDKRGANGARLALMPQRDWDVNAAAVRAALPVLEKIQ 517  
QY 611 ASDSRFLLSWLEQARAAAVSEA 633  
DB 518 KESGKASLADIIVLAGVVGVEKA 540

Search completed: September 21, 2002, 12:24:27  
Job time: 42 sec



Db	361	WGPAQIRAVLGVPRGRLLVLDLFAESQPVYTRTASFQGPPIWCMHLNFCGNHGLFGAL	420	
Qy	421	EAVNGSGPEAARLFPNSTVMGTGMAPGEGISQNEVYSLMAELGWRKDPDPDLAAWVTSFAA	480	
Db	421	EAVNGSGPEAARLFPNSTVMGTGMAPGEGISQNEVYSLMAELGWRKDPDPDLAAWVTSFAA	480	
Qy	481	RRYGSHPDAGAAWRLRLLSVYNCSEACRGNHRSPLVRRPSLQWNTSIWYNRSVDVEAW	540	
Db	481	RRYGSHPDAGAAWRLRLLSVYNCSEACRGNHRSPLVRRPSLQWNTSIWYNRSVDVEAW	540	
Qy	541	RLLLTSAPSLATSPAFRYDLDLTQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY	600	
Db	541	RLLLTSAPSLATSPAFRYDLDLTQAVQELVSLYEEARSAYLSKELASLLRAGGVLAY	600	
Qy	601	ELLPALDEVLASDSRLLGSLWEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN	660	
Db	601	ELLPALDEVLASDSRLLGSLWEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN	660	
Qy	661	KQLAGLVANYTPRWRLFLLEALVDSVAQGIPIFQOHQFKNVFLQEQAFVLSKORYPSQPR	720	
Db	661	KQLAGLVANYTPRWRLFLLEALVDSVAQGIPIFQOHQFKNVFLQEQAFVLSKORYPSQPR	720	
Qy	721	GDTVDLAKKIFLKYYPGWAGSW	743	
Db	721	GDTVDLAKKIFLKYYPGWAGSW	743	
RESULT 2				
C87316				
alpha-N-acetylglucosaminidase [Imported] - Caulobacter crescentus				
C:Species: Caulobacter crescentus				
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001				
C:Accession: C87316				
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.				
n, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon				
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001				
A:Title: Complete Genome Sequence of Caulobacter crescentus.				
A:Reference number: A87249; PMID:21173698; PMID:11259647				
A:Accession: C87316				
A:Status: preliminary				
A:Molecule type: DNA				
A:Residues: 1-770 <STO>				
A:Cross-references: GB:AE005673; NID:gl3421729; PIDN:AAK22527.1; GSPDB:GN00148				
C:Genetics:				
A:Gene: CC0540				
C:Superfamily: human alpha-N-acetylglucosaminidase				
Query Match 26.9%; Score 1060.5; DB 2; Length 770;				
Best Local Similarity 33.0%; Pred. No. 1.3e-68;				
Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;				
Qy	2	EAVAVAAVGVLLLAGAGGAAGDEARAAVRAVRLVLLGPGPAADFVSVERALAAKPG	61	
Db	11	QATSLAAAFVAPALAAAGTDG----VAAARASLKRFLGRRLAG-----AHLTVTPG	59	
Qy	62	LDT--YSLGGGAARVRVRGSGTVAAAGLHRYLRDFCGCHVAGSWGSQLRPLRPAPVG	119	
Db	60	AERSWYAIGKGA--ISIGSDSPVALVRGAYAHLRQAGLAHVSWEGRVQAGVAPAGAG	118	
Qy	120	ELTEATPNRYVYQVNTQSYFVWDMARWERIDMNLNGINLALAWSGQEAIVORVY	179	
Db	119	ARVE--TFPRHRAVYLVNTCYGTYTTPWGWGRWTRIEDIDWAAHGIDMPLAMEGQEVVNRALW	177	
Qy	180	LALGLTOAINEFTGPAFLAWGBMGNLHTWDGPLPSPWHIKOLYLQHRVLDQWRSFGMT	239	
Db	178	REFGLSBAELADYFSGPAFTFWHRNGNIEGKAPLPTAWIDKKDLQVKILGRNRSLSMT	237	
Qy	240	PVLPAFAGHVPEAVTRVFPVNTVMKMSGWGHFNCYSFCSFLLAPEDPIFFPIIGSLFUREL	299	
Db	238	PILPAFGGYVPKAFAEKNPKARIYMRPWEGFHETY----WLDPADPLFAKIAARFLALY	293	
Query Match 3.1%; Score 123.5; DB 1; Length 497;				
Best Local Similarity 22.9%; Pred. No. 0.45;				
Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;				
Qy	343	TEAWLL-----OGWLFQHQPFQWGAQIRAVLGVPRGRLLVLDLFAESQPVYTRT	394	
Db	12	TVAIFLLVLDLHRRQRWAARYPP---GPLPLPGL-----GNLLHVD-----	50	
Qy	395	ASFQGGPFTWCMHLNFCGN-----HGLFGALEA-VNGGPEAARL--FPNST	437	
Db	51	--FRNTPYCEDQLRRRFGNVSFLQLAWTVPVVLNGLAAVREALVTCGEDTADRPPIQ	108	
Qy	438	MVGTGMAPEGI-----SQNEVYSLMAELGWRKDPDPDLAAWVTSFAARRVGVSH	487	

```
Db 109 VLGFGRSQGVFLARYGPANREQRRFSVSLRNILGLK---KLEQWVTEEARCLCAAT 165
QY 488 PDGAARW---LLLSRVYNSGACRGHNRSPVRRPSLOMNTSIIWYNSRSDVFEARWLL 544
Db 166 DQAGRFRPNSLDKAVSN-----VIASLTYGR----- 193
QY 545 TSAPSLATSPAFRYD-----LLDLTRQAVQELVSLYYEARSAYLSKELAS---LLR-- 593
Db 194 -----RFEYDDPRFLRFLDLTHEALKE-----ESGFUREVINAIPLLLRIP 234
QY 594 --AGGVLAYE--LLPALDEVL-----ADSRELLGSWLEQARAAAVSEAEADFYEQ 640
Db 235 GLAGKVLRSOKAFLTQDELLEHRTMTDPAQPPRDLTEAFLAEMEKAR-GNPESSENEE 293
QY 641 NSRYQLTLMGPEGNILDYANKOLAGLVANVYTPRWRLFLLEALVDSVAQGIPIQQOHFDKN 700
Db 294 NLR-----MVVADLSAGNVTTSTLAWGLLLMLHPDVQRRV---QQEIDDV 338
QY 701 VFQLEQAFLVSKORYP-----SQPRGDTVDL 726
Db 339 IGQVRRPEMGDQARMPYTTAVIHEVQRFGDIVPL 372

RESULT 4
H83343
Probable non-ribosomal peptide synthetase PA2424 [Imported] - Pseudomonas aeruginosa (st
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83343
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: H83343
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4342 <STO>
A:Cross-references: GB:AE004669; GB:AE004091; NID:99948460; PIDN:AAG05812.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2424
C:Superfamily: acyl carrier protein homology; acetate--CoA ligase homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:60-553/Domain: acetate--CoA ligase homology <ACLI>
F:584-652/Domain: acyl carrier protein homology <ACP1>
F:1174-1622/Domain: acetate--CoA ligase homology <ACLI2>
F:1637-1705/Domain: acyl carrier protein homology <ACLI2>
F:2232-2689/Domain: acetate--CoA ligase homology <ACP2>
F:2706-2773/Domain: acetate--CoA ligase homology <ACLI>
F:3759-4230/Domain: acyl carrier protein homology <ACP3>
F:4248-4316/Domain: acetate--CoA ligase homology <ACLI4>
F:1669,2738,4280/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.18; Score 121; DB 2; Length 4342;
Best Local Similarity 21.4%; Pred. No. 14;
Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;

QY 4 VAVAA-----AVGVLLLAGAGGA-----AGDEARCAAAVRA-LVARLL 40
Db 1181 VAICERSPQLLVGLLAIVRAGGAYVPLDDPYSERLAYMLADSGVELLLTQAHFLERL- 1239
QY 41 GPG-----PAADFVSVERALAAKPGLDYSLGGGGAAVRVRGSGTGVAAAAGLHRYLRD 95
Db 1240 -PGAEGVTPICLDSKLDNNPSPQAGLHLH--GDNLAIVYITSGSTQPKGVG----- 1289
QY 96 FCGCHVAVSGSOLRLPRLPFAVGE--LTATPNRYRYQNVCTQSYFVWMDWARWERE 153
Db 1290 --NTHAALA-ERLQWMOATYTLGDGDDVLMQAP-----VSFDVSVW--E 1328
```

```
QY 154 IDWMLANGINLALWSGQ-----EAIWQVYLLALGLTOAEINEFFTGPAFLAWRGM 205
Db 1329 CFWPLVTGCRVLAAAPGEHRDPARLVELVRQFGVTTLHFVPPLLQLFIDEPGVAACGSIR 1388
QY 206 NLHTWDGGLPPSPSHLKLQLYLQHRVLDQWRSFGMTVPVLPFAFAGHVPVATRVPPQNVTKM 265
Db 1389 RLFSGEALPAE-----LRNRVLR-----LPAVALHNRYGTET--AINVT-- 1428
QY 266 GSWGHFNGSYSCSFLAPED-----PIFFIIGSLFLRELKKEFGTDHIYGADTFNEM--- 317
Db 1429 -----HWQC-----RAEDGERSPIGRPLGNVVCVLDAAEF---NLLPAGVAGELCIGG 1473
QY 318 -----QPPSSPSPLAAATTAAYEAMTAVDTFVALLQGWL-----FQHQPFQWG- 362
Db 1474 LGLARGYLGRFALSARFVADPFSAAGRLYRTGDRARNADGVLEYLGRLDQOVKLRGF 1533
QY 363 ---PAQIRAVILGAVP---RGRLLVLDLFAESQPV---YFRTASFQGFQPFIMCLHNFNGNH 414
Db 1534 RIEPEIOARLLAOPGVAQAVVIREGVAGSOLGVYTYTGAVGAAEA-----EQNQ 1584
QY 415 GLFGALEAVNGGPEAARLFPNSTMVGTGMAPEGISQNEVVYSLMAELGWRK---DPVPD 470
Db 1585 RLRAALQA-----ELPEYMWPTQLMRLAQMPLGPS-GKLDTRALPEPVWQQRHEHVRTE 1638
QY 471 L-----AAWVTSFAARYG-----VSHPDAGAAWRLLLRSVYNC-- 505
Db 1639 LQRIIAIWSVGLPRVGLRDDDFELGGHSLLATRIVSRTRQACDVLPURLFAESEL 1698
QY 506 -----GEACRGHNRSPVLRPSLQMNSTSIWNKSDVFEAWRLLLTSAPSLATSPAF 556
Db 1699 EAFCEQVRAAAGRTDSHGAIIRDREOPVPLSYSQRMWFLMQL-----EPD--SPAY 1751
QY 557 RY-----DLLDLTR--QAVOELVSLYYEARSAYLSKELASLLRA---GGV-LAYELL 603
Db 1752 NVGGLARLSPGLDVARFEAALQALVQ-RHETLRTFFSVDGVPQVRVHGDLHMDQDF 1810
QY 604 PALD-----EVLASDRFLGLSLEQA---RAAAVSEAEADFYEONSRYQLTLWGPE 652
Db 1811 SALDRSQRQHLQFLA-DSEAHRFDESLGPLLRVCVMKWAERHYLVTLHHIV---TE 1866
QY 553 GNILDYANKQLAGLVANVYTPRWRLFLLEALVDSVAQGIPIQQOHFDKNVQLE 705
Db 1867 GWAMDFARELGALYEAFLLDDR-----ESPLELPVQ--YLDYSVWQRE 1908
```

RESULT 5  
S18268

delta-(L-alpha-aminoadipyl)-L-cysteinyI-D-valine synthetase - Streptomyces lactamdura  
C:Species: Streptomyces lactamdurans  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Nov-2000  
C:Accession: S18268; S15283; B38171  
R:Martin, J.F.  
submitted to the EMBL Data Library, January 1991  
A:Reference number: S18268  
A:Accession: S18268  
A:Molecule type: DNA  
A:Residues: 1-3649 <WAR>  
A:Cross-references: EMBL:X57310; NID:g45005; PIDN:CAA40561.1; PID:g45006  
R:Coque, J.J.R.; Martin, J.F.; Calzada, J.G.; Liras, P.  
Mol. Microbiol. 5, 1125-1133, 1991  
A:Title: The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptid  
genes in Acromonium chrysogenum and penicillium chrysogenum.  
A:Reference number: S15283; MUID:92065808  
A:Accession: S15283  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 224-940; 1319-2010; 2373-3307 <COQ>  
A:Cross-references: EMBL:X57310  
A>Note: the source is designated as Nocardia lactamdurans  
R:Coque, J.J.R.; Liras, P.; Laiz, L.; Martin, J.F.  
J. Bacteriol. 173, 6258-6264, 1991  
A:Title: A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precu  
A:Reference number: A38171; MUID:92011390

A;Accession: B38171  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-23 <CO2>  
A;Cross-references: GB:S57006  
C;Genetics:  
A;Gene: pcbAB

A: Gene; C: Superfamily; alpha-aminoacyl-pyl-cysteine-lyl-valine synthetase; acetate--CoA ligase homology  
C: Superfamily; cephamycin biosynthesis; phosphopantetheine; phosphoprotein  
C: Keywords; carrier protein; acetate--CoA ligase homology <ACLI>  
F: 298-758/Domain: acetate--CoA ligase homology <ACLI>  
F: 786-856/Domain: acyl carrier protein homology <ACP1>  
F: 1392-1844/Domain: acetate--CoA ligase homology <ACL2>  
F: 1862-1932/Domain: acyl carrier protein homology <ACP2>  
F: 2446-2895/Domain: acetate--CoA ligase homology <ACL3>  
F: 2912-2980/Domain: acyl carrier protein homology <ACP3>  
E: 202-1896 2944/binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.0%; Score 119; DB 1; Length 3649;  
Best Local Similarity 21.1%; Pred. No. 16;  
Matches 174; Conservative 63; Mismatches 283; Indels 306; Gaps 37;

Qy	16	AGAGGAGDEAREAAARLV-----ARLLGPGPAADFVSVERALAAKPGDLTYSLGG	69
ph	1098	AAI DCAICCTIVVDDPAIPTIKTDGQGVGRQVPTIANDVRIEVP-----2031	2031

120

*[The following section contains faint, illegible markings or bleed-through from another page.]*

----- 2302 00

QY 130 RYYQNVCTQSYSFVWWDWARWEREL-----DWMALNGINLALAWSGGEATWQKVI 173

db 2071 EIYLSVVVHHSCFDGWSWDIFRRRELAALLDGVPEADLGALRGT-----YGEFAVWQRQY 2124

QY 180 LALGLTQAEINEFFTGPFLAWGRMGNLHT---WDGPLPPSWHIK---QLYLQHRVLD 231

Db 2125 • LT-GKRLAALTEFWT-----GALGGFETIALPDHPRPPFDYRGRELEFELDERTE 2176

OV	232	QMRSEGMT	-----	239
OV	232	QMRSEGMT	-----	239

2177 A:BELARTARVSI.YSVI.I.GAWCLMLNMYTGOHDLVVGTPSANRGRPEEFDRAVGFFANLLA 2236

C: 240 ----- DWT DAFACUVDENVTBIVEOQVNVTKMGSSWGHENCYSQCSFELAPEDPT-FPT 290

[illegible][illegible]

QY 291 1GSEFKREFGDRTIGADIFNEMQTCSBIOEUAHIVTETZKRVSTERRWEE

DB 2288 LQLNFTLQNV-----SDH---TSALTGYQPDS-----GGWTTTKFD-LSAIMIEIAIGLA 2333

Qy 351 GWL-----FQHQPWFVGPAQIRAVLGAVPRGRLVLDLFAESQ-PVY 391

Db 2334 GNLTAAASLFDDTSASGFIATFKHV-----LAEFASAAQTPIAQLTALDEPGQAALPDA 2388

OV 392 TRTASFOGQPIWCMLHNFGGNHGLFGALEAVNGGPEARLFPNSTMVGTG--MAPEGI 448

[illegible]

**THE UNIVERSITY OF CHICAGO**

[illegible]

QY 494 W-----RLLRSVYNCSGEACRGHNRSPLVRRPSLQMN'TSLWYNKRSDFEAWRLLEIS 348

Db 2491 FMLS DTGAKLVL-----AGEAHGSRVRG-----LTSGDVL DLEQLDLTG 2529

QY 547 APSLATSPAIFYDLDLTRQAVQELVSLYYEARSAYLSKELASLLRAGGVLAYELPAL 606

Db 2530 EP--AENP-----VTETSTELAYAIY--TSGTTGKPKAVLVSHGSV-----2567

Q: 607 DEVI ASDSPEILCSWLEOARAAAVSEAFADFEQNSRYOITIWGPEGNLT----- 656

[illegible]

1

```
QY 657 ----DYANKOLAGLVANYYTPWRFLFLEALVDSVAQGIPFQOHOED 698
      : ||: || | | | | | | | | | |
Db 2625 PAFYELANRE--GL-----SYLSGTPTOVERFD 2650
```

## RESULT 6

O4HUD1  
debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome p450 2D6 - human  
N: Alternate names: CYP2D6; cytochrome p450 isozyme 2D; cytochrome p450db1  
C: Species: Homo sapiens (man)  
C: Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 03-Mar-2000  
C: Accession: S01199 #accession: J04156; A33629; A30335  
R: Gonzalez, F. J.; Skoda, R. C.; Kimura, S.; Umeno, M.; Zanger, U. M.; Nebert, D. W.; Gel  
Nature 331, 442-446, 1988  
A: Title: Characterization of the common genetic defect in humans deficient in debriso  
A: Reference number: S01199; MUID:88122614

A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-497 <GON>  
A:Cross-references: EMBL:X08006; NID:g30450; PTDN:CAA30807.1; PID:g30451  
R:Gonzalez, F.J.; Vilbois, F.; Hardwick, J.P.; McBrick, O.W.; Nebert, D.W.; Gelboin, Genomics 2, 174-179, 1988  
A:Title: Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid sequence  
A:Reference number: A28883; MUID:88314109  
A:Accession: A28883  
A:Molecule type: mRNA  
A:Residues: 1-497 <GON>  
A:Cross-references: EMBL:M20403; NID:g181349; PTDN:AAA52153.1; PID:g181350  
R:Jiang, Q.; Voigt, J.M.; Colby, H.D. Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995  
A:Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16)  
A:Reference number: JC4153; MUID:95251703  
A:Accession: JC4156  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-497 <JTA>  
R:Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J. Am. J. Hum. Genet. 45, 989-904, 1989  
A:Title: The human debrisoquine 4-hydroxylase (cyp2d) locus: sequence and identification  
A:Reference number: A33629; MUID:90072069

A. Residue: 1-373, 'V', 375-497 <IM>  
A. Residues: 1-373, 'V', 375-497 <IM>  
A. Cross-references: EMBL: M3388; NID: g181303; PTDN: AAA53500.1; PID: g181304  
R. Manns, M.P.; Johnson, E.F.; Griffin, K.J.; Tan, E.M.; Sullivan, K.F.  
J. Clin. Invest. 83, 1066-1072, 1989  
A. Title: Major antigen of liver kidney microsomal autoantibodies in idiopathic autoimmune hepatitis  
A. Reference number: A30335; MUID: 89155788

A: Molecule type: mRNA  
A: Residues: 125-73, 'V', 375-485, 'T', 487-497 <MAN>  
A: Cross-references: EMBL:M24499; NID:G522194; PIDN:AAA36403.1; PID:G522195  
C: Genetics:  
A: Gene: GDB:CYP2D6  
A: Cross-references: GDB:132127; OMIM:124030  
A: Cross-references: 22q12.1-22q13.1  
A: Cross-references: 22q12.1-22q13.1

C:Superfamily: human cytochrome P450 CYP2B6; Cytochrome P-450 hemase; iron; metal binding site; heme iron (cys)  
E:302-465/domain: cytochrome P450 homology <CYP>  
F:442/motif: heme iron (cys) (axial ligand) #status predicted

Query Match	3.0%	Score 117.5;	DB 1;	Length 497;
Best Local Similarity	21.7%;	Pred. No. 1.2;		
Waterbase	97. Conservative	53. Mismatches	144. Indels	153. Gaps
				22;

345 AVVLL-----QGWLFHQHQPFWGPPQAIRAVLGVAPRGELLVLDLFAESQPVYTRTAS 396

Query Match	3.08;	Score 117;	DB 2;	Length 1374;	
Best Local Similarity	20.08;	Pred. No. 5.6;			
Matches 163;	Conservative 78;	Mismatches 249;	Indels 324;	Gaps 37;	
QY	1	MEAVAVAAAGVILLAGAGG-----AAGDEARENAAVRA-----	LVAR 38		
Db	620	IDSITLETGVPHFHTAAGNGSGPLGTVPATAPARLAVAVAAATDMAYLSLIQPGYLPLLAG	679		
QY	39	LLGPGPAADFVS-SVERALAAKPLDTSYLGSGGAARVVRGSGTCVAAAAGLHRYLRDFC	97		
Db	680	LGGYGDPAYFSARGSPHAGKPGI-----AAIGCFAYTTGRS	717		
QY	98	GCHVAVSGSQLRPLRPAPVPGELTEATPNRYRYQNVCTQSYFVMDWARWEREIDMW	157		
Db	718	LDH--YTGGRLD--PRAAPLLFGGTSMATP-----M	744		
QY	158	ALNGINIAL----AWSQGEAT----WQRYVALGLTOAEINEFTTGPAPFLAWGRMNLHT	209		
Db	745	AAGAAALAIQAKESLGVRLGLEWLRVYALTSMTAQ-----WRG---LPWGEWGN---	793		
QY	210	WDGPLPPSWHTK-----OLYLQHRVLDQMR-----SFGM-----TPV	241		
Db	794	--GIVDAAGAIRLTLTGVDQGVLIYSATILIEAAGQAVAAPGYPGIPALLWAGSGVETPV	851		

Query Match 2.9%; Score 115.5; DB 2; Length 973;  
Best Local Similarity 18.4%; Pred. NO. 4.4;  
Matches 165; Conservative 93; Mismatches 293; Indels 347; Gaps 45;

Qy	1	MEAVAAAGVLLAG-----AGGAAGDEAAAAARVALRLGLGPPAADFSV	55
Db	225	MGASADTPGVTVAAAGPTTVVASPGQLGVQGESGTYEV---KLSQPTANVTVT	280
Qy	56	LAAPGDLTYSLLGGG-----AARVV-----RGSTGVAAGLRHRYLRDFCG	98
Db	281	-SGNTGL---TLTGASLTTPSNWDTAQRVTVSADASGSGSAVFESTAPGHG-	329
Qy	99	CHVAWSGQLRLPLPA---VPGELTEA-----TPNRYRY-----QNVK	136
Db	330	-KAATVTVLAAAKDYDARELELYGKITDPANGYFSGEPIPHYSVETLIVEAPD	388





Db 52 QAYLQHRDLPAVAREGWI-----AKAAEIMERDITAKFADVLVDEIGSGDIAKAGETPRAPS 107  
QY 279 FLA-----PEDIPPIIGSLFLRELKEFGDTHIYGADTFENEMQPPSSEPS 325  
Db 108 FLRAAIGVPRIRGETIPSDT--PGRFMSLRQPV-----GVVAGITPFNVPLIKGIQS 160  
QY 326 YLAATAATVAYEAMTAVDEAVMLLQGLFQHQFQNGPAQIRAVLGNVPRGRLLVLDLFA 385  
Db 161 AMALATG--NAFVLLPSEAPMIADLL-----AKLKEA-----GVDPGLFNV----- 201  
QY 386 ESQPVYTRTASFOGQPIWCMHNFNGHGLFGALEAVNGPEAARL-FPNSTMVGTGMA 444  
Db 202 ----VY-----GNGAEIG--DVLTGHPKVASITFTGSSRVCKHTA 235  
QY 445 PEGISONEVYVSMALGWRKQVPDAAWVTSFAARRVGVSPDAGAAWRLLRLRSVYNC 504  
Db 236 -ETAARNLKTYL--ELG-GKSLVICA-----DADLDKAVNAALFSIFMY 277  
QY 505 SGACRCHNRSPVRRPSLOMNTSIWNRSDVPEAWRLRLITSAPSLATSPAPRYDLDDLT 564  
Db 278 QGVQCMGASR-----IVYERS-IFDQF-----TKAFAAATGRANSGLDRDPT 318  
QY 565 -----ROAVQELVSLVYEEASAYLSKELASLLRAGGVLAYELLPALDEVLADSREL 617  
Db 319 TMLGPIIISERQRRVRHDDAKS-----KGAAVLAG----- 350  
QY 618 LGSWLEQARAAAYSE---AEADYEQNSRYQLTLWMEGNI---LDYANKOLAGLVANY 671  
Db 351 -GEWSGNSCAATILSGVTAEMTVEEETGPTVSLFPFDTLBEALELANNTEYGLSASIF 409  
QY 672 T 672  
Db 410 T 410

RESULT 11  
hypoetical protein PA1923 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83405  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83405  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1281 <STO>  
A:Cross-references: GB:AE004618; GB:AE004091; NID:g9947912; PIDN:AAG05311.1; GSPDB:GN001  
A:Experimental source: strain PA01  
A:Genetics:  
A:Gene: PA1923

Query Match 2.9%; Score 113; DB 2; Length 1281;  
Best Local Similarity 22.5%; Pred. No. 9.8;  
Matches 104; Conservative 44; Mismatches 141; Indels 174; Gaps 26;

QY 349 LOGWLFHQPOFWGPAQIRAVLGAVP-----RGRLLVLDLFAESQPVYTTA---SFOG 399  
Db 40 LAGWAREAGVELRG---LRLGIGEAPCEWLDDGNLLILDTPRTPDRAQVEEALGERLQG 96  
QY 400 --OPFIWMLHNFNGHGLFGALEA-----VNGGP-----EAARLFPNSTVMVG 440  
Db 97 GTQPMW-----RVGGPGFGNLPALGGRVLGVYANGGEANRLFEAVRRWHAGLTVD 151  
QY 441 TGMAGEISQ-----NEVYISLMAE--LGRKDPVFDLAAWVT-----SFAARRYGVSH 487  
Db 152 ALPAPQPLAQAGFYHPDAPFAVLADYLA-----GASRWASDAPRIAPLIPRGAIA 205

QY 488 PDAGAAWRLLRSVYNCSEACRGNRSPLVRRPSLOMNTSIWNRSDVFEAWR----- 541  
Db 206 AQTGAIDELLRS-----ERHQAPL-----AVFDDSDP-EALRKSFGA 245  
QY 542 --LLTSAPSLATSPAFRYDLDLTRQAVOELVSLYEE-----ARSAVLSKELAS 591  
Db 246 DVQALVNLQHLONGPARRAEFLALD---VPVLTQLCYRDGNEADWLAAASGVAPRTAAAF 302  
QY 592 -----LRAGGVLAYELLPALDEVL-----ASDSRELLGSW 621  
Db 303 LGMPTWGSDDPLVISALENGEPKLMAG--QAEALLDKLDRLLRLRPAADKHLALMFW 360  
QY 622 L-----BQARAAVSEAEADFEQNSRYQLTLWMEGNI 658  
Db 361 NHPEGEKNVAASHLNVPSLARLGEALRAAGYRVATSD-----ESALDIT 405  
QY 659 ANKQLAGLVANYTYRWRFLFLEALV-DSVAQGIPEQQ--HOFD 698  
Db 406 AORLLGG---YIRPQ---TUDALYRDGLAASLPLDAYLHWFE 441

RESULT 12  
I78557  
N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 21-Jan-2000  
C:Accession: I78557; I58158; D45219  
R:Monyer, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.  
Neuron 12, 529-540, 1994  
A:Title: Developmental and regional expression in the rat brain and functional properties of the N-methyl-D-aspartate receptor subunits NR1 and NR2B  
A:Reference number: I58158; MUID:94206533  
A:Accession: I78557  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1323 <RES>  
A:Cross-references: GB:L31612; NID:g469068; PIDN:AAC37647.1; PID:g469069  
A:Accession: I58158  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-66, 'V', 68-1323 <RE2>  
A:Cross-references: GB:L31611; NID:g469066; PIDN:AAC37646.1; PID:g469067  
R:Ishii, T.; Moriyoshi, K.; Sugihara, H.; Sakurada, K.; Kadotani, H.; Yokoi, M.; Akazawa, J.  
J. Biol. Chem. 268, 2836-2843, 1993  
A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits  
A:Reference number: A45219; MUID:93155102  
A:Accession: D45219  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1265-1323 <ISH>  
A:Experimental source: brain  
A:Note: sequence extracted from NCBI backbone (NCBIP:124265)  
C:Superfamily: N-methyl-D-aspartate receptor 2b; glutamate receptor homology F; 451-879/Domain: glutamate receptor homology <GRH>

Query Match 2.8%; Score 112; DB 2; Length 1323;  
Best Local Similarity 22.9%; Pred. No. 12;  
Matches 106; Conservative 48; Mismatches 157; Indels 152; Gaps 28;

QY 162 INLALWSGQEAIWORYVLALGLTQAEINEFFTGPAFLWGRMGNLH-----TWGGLP 215  
Db 48 LNLVAFVSGP-----AYAAEAARL-----GPAVAAAVRSPGLDVRPVALVNGSDP 93  
QY 216 PSWHIKOLYLIQ-----HRVL--DQMRSGMPTPLPAFAGHVEATRVFPQVNT 263  
Db 94 PS-----LVQLCDLLSGLRVHGVFEDDSRAPAVAPILDFLSAQTSPLIVAVHG----- 143  
QY 264 KMSGWHFNCYSFCLAPEDPIFFIIGSLFLRELKEFGTTHIYGADTFENEMQ----- 318  
Db 144 -----GAALVLTPEK-----GSTFLQ-----LGSSTEQQLQVIFEV 175  
QY 319 -PPSSPSPYLAATTA-----VYEAMTAVDTEAVMLQGLWLFQHQPOF--WGAQIRAVLGA 372

Db 176 LBEYDWTSEVATTPRACHRAFLSVIEVLTGD--SLVGW--EHRGALTDPGAGEAVLGA 231  
Qy 373 VPRG-----RLVLDFAESQPVY--TRTASFQGGPFTWML-HNFGNHLGFGALEAVN 424  
Db 232 QLRYSVAQIRLLFC-AREAEVFRABEAGLTGPGYWMFVGPQLAGGG-----S 282  
Qy 425 GQPEARLPNPNWGTGMAPGIGSQNEVYSLMAELGWRKDPVPLAAWVTSFA----- 479  
Db 283 GYPGEPLLLPGGSPPLAGL-----FAVASAGRODLARVAVGAVVARGAQA 330  
Qy 480 -ARRYGVSHPDAGAAWRLRLRSVYNSGEGACRGNRSPLVRPSLQ---MNTSIWYNSD 535  
Db 331 LLRDVGLF-PEL-----GHDCRTQNRTH--RGESLHRYFMNIT-WDNRDY 371  
Qy 536 VFEAWRLRLTSPASLATSAPFYDLDTLQRAVQELVSLYEE 578  
Db 372 SFNEDGFLVN--PSLV-----VISLTRDRTWEEVGSWEQO 404

RESULT 13  
H87316  
hypothetical protein CC0545 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: H87316  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: H87316  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1027 <STO>  
A:Cross-references: GB:AE005673; NID:g13421736; PIDN:AAK22532.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0545

Query Match 2.8%; Score 111; DB 2; Length 1027;  
Best Local Similarity 22.5%; Pred. No. 10;  
Matches 157; Conservative 47; Mismatches 217; Indels 278; Gaps 38;

Qy 135 VCTOSYSFVN-----WDWARWERE-----IDWMLNGINLAL-----AWSQGEATW 175  
Db 68 VCDVSATGQWLLRASERAWDLKRLNDEGSLRKAIGWRRLDGADPATADLLDALLGKGAFS 127  
Qy 176 QRVYALGLTQAEINEFTTGPFLANGRMGNLHTWDGPIPPSWH-IKQLYLQHRVLDQMR 234  
Db 128 PKA-MARMLERIEAQEAAGPEL---RLTATALEWSGALAPAAALLARQAALNRRTDOAR 183  
Qy 235 SFGMTPVLPAPAGHVPEAVTRVPPVNVTKMSGHFNCSYSCSFLIAP-EDPI----- 287  
Db 184 R-GELLADGFGQR-----EQAVEILTW-----IRNPODDPIRVRLY 221  
Qy 288 ---FPPIIG-SLFLRELKIFGTCHI-----YGATFNEM-----OPSPSEP 324  
Db 222 VSGLPAGKSTLLEHLVLRQAQGVETVVVRLDFDRASLDVLDQLGLSVEVARGVASQLP 281  
Qy 325 SYLAATATYVEMT-----AVDTEAVWL 348  
Db 282 SAAADLRQARLEAITTTQETIRLKGGRSGTGYALIQALGVLQAGSAGRLRVWVLDLTLEVL 341  
Qy 349 LOGWLFQHQPOFWGPAQITRAVL-----GAVPRGRLL--VLDFLFAES----- 387  
Db 342 GRGETHPQLDFDLQDLARVLVPLVVVAGRGRALEPVDRAVETLNLDRLTDAEERL 401  
Qy 388 -----QPVYTRT-ASFQGGPFTWMLHNFNGNHG-----LFGALEAVNGG 426  
Db 402 LSAADVPPALMTQVIALAEGDPLQLRLATLARDEQAGLTGKGATITLYRSLEARLGG 461  
Qy 427 PEA--ARLFP-----NSTMVTGMGP-----EGISQNEVYSLMAEL-----GW 463

Db 462 RHAKLARFIPLFRFENSELLGAVMAPVVLGERLSPGAAA-ALMAELKDOTWLIAGTEGW 520  
Qy 464 RKDPVPD-----LAAWVT'SFAARYGVSHPDAGAAWRLRLRSVYNSGEGACRGNRSPLVR 519  
Db 521 -LTPQDRLRRRLACKLYDEHARRTGALHRRR-ARK-----FAE 556  
Qy 520 RPSLQMNITSIWYNSDVFEAWRLRLTSA---PSLATSAPFR-YDLLDLTROAVQELVSLY 575  
Db 557 RPELWAGAESLYHR---LOATRWAGPDALKGMNLAAGVAFQASADLELP----- 602  
Qy 576 YEEARSAYLSKELASLRAGGVLAYE-----LLP-----ALDEV-----LASDSR 615  
Db 603 -EABDA-----LLRARGERSYEGRDALTPSGEVSPAAADELRMLTSGDGLAEASH 653  
Qy 616 FL-----LGSWLEQARAAA 629  
Db 654 FYGRVFRPAALDPTSPADVALTLTLWRTGRTWEARRLAA 692

RESULT 14  
F87285  
conserved hypothetical protein CC0295 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: F87285  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: F87285  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <STO>  
A:Cross-references: GB:AE005673; NID:g13421436; PIDN:AAK22282.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0295

Query Match 2.8%; Score 109.5; DB 2; Length 285;  
Best Local Similarity 23.7%; Pred. No. 2.1;  
Matches 80; Conservative 36; Mismatches 90; Indels 131; Gaps 17;

Qy 328 AAATTAAYEAMTAVDTAEAVMLQGLFQHQPFQWGPQAIQAVLGVAPRGLLVLDFAES 387  
Db 3 AAPTAWAGLTAYSWLATTICAAWF-----ARRFALADGRVLSVE----- 43  
Qy 388 QPVYTRTASFQGPFE-----TWCMLHNFNGNHLFGALEA----- 422  
Db 44 -----OSLAQQGAIYAAWLPVIGIWLVLRRFGA--GLRGVLAATPATGLVIVPLEALVAS 96  
Qy 423 -----VNGSGEAPAR-----LFPNSTWVTGTM-----APEGISQNEVYSLMAE- 460  
Db 97 LIDOTEFAGGDLAERALGRAPVCILLHTAIVAGVLAHRAAASARNTLLQAALAEA 156  
Qy 461 -----LGRKDPVPDLAAWVTSFAARYGVSHPDAGAAWRLRLRSVYNS 505  
Db 157 RATSVLDVAERLMMVAGARRVPV-DITAVENFGAANDYVVVHWDGREG---LMRATLQ-S 211  
Qy 506 GEA-----CRGNRSPLVRPSLQMNITSIWYNSDVFEA-----WRLLTSPASLA 551  
Db 212 LEARLDPLRFARAH-RSALV-----NLSKVSQAQPLSDGNSRLTLASGAEVV 257  
Qy 552 TSPAFRYDLDD-ITROAVQELVSLYEEARSAYLSKE 587  
Db 258 TSRTYRDALLKRLGROA-----SDSQASMSSE 285

RESULT 15  
T47641  
hypothetical protein T15C9.20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47641  
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
Submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224470  
A:Accession: T47641  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-814 <MEW>  
A:Cross-references: EMBL:AL132970  
A:Experimental source: cultivar Columbia; BAC clone T15C9  
C:Genetics:  
A:Map position: 3  
A:Introns: 28/2; 51/3; 252/3; 312/3; 350/3; 370/2; 392/3; 410/1; 441/1; 469/1; 604/3; 62  
A:Note: T15C9.20

```
Query Match          2.8%; Score 109.5; DB 2; Length 814;
Best Local Similarity 20.6%; Pred. No. 9.3;
Matches 113; Conservative 69; Mismatches 192; Indels 175; Gaps 26;

QY 150 WEREIDWALNGINLAWSGOEAIWQRYVYALGLTQAEINEFFETGPAFLAWGMGNLHT 209
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 WKEEVLIRGGVPMAL-----RGELWQAF---VGVKRRCKDYQ-----NLLA 274

QY 210 WDGPLPPSWHIKQLYLQHRVLDQMRSCGMTVPVPAFAGHVPFAVTRVPPQVNVTKMGSMG 269
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 ADGSVNT---IEQEDMQH-VDDKGSSTESIAVVEKWKQIEKDLPRFP-----G 320

QY 270 HFNCSYSCSFLIAPEDPIFFIIGSLFRELKEFGTDHIYGADTFNEMQPPSPSYLAA 329
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 H-----PALDDGGRNALRRLLTAYARHN-----PSVGYC 349

QY 330 ATTAVYEAMTAV---DTEAVVLLQGLWFOHQPQWPAQIRAVLGAVPRGLLVLDLFAE 386
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 QAMNFFAALLLLMPENAFWALGLIDDDYNGYIYSEEMIESQV-----DQLVLEELVRE 404

QY 387 SOPVYTRTASFGQGFIFWCMILHNFNGNHGLFGALEAVNGSGPEAARLFPNSTMVGTMGP- 445
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 RFPKLVHHLVDYLGQVAVVT-----GPFELSIEMN-----MLPW 438

QY 446 EGISQNEVVYSILMAELGWRKDPVPLAAWVTSFAARR-YG---VSHPDAGAAWRL---LL 498
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 439 ESVLR---VMDVLLFEGTR-----VMLFRTALALMELYPALVTTKDAGDAVTLLOSLT 489

QY 499 RSVYNCSG---EACGRHNRSPVRRPSLQMN-----TSIWNRSDDVFEAWR----- 541
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 GSTFDSSQVLTCMGYQNVHVEIRLQELRSKRPVAVTALEERSKGLQAWRDSKGLASKL 549

QY 542 -----LLLTSAPLATSPAFRYD-----LLDLT-----RQAVQELVS 573
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 YNFKQDPKSVLVDKSLNGSLSGSSGSSNADEVLSVLTGDEVDVSVQDLQAQAEICK 609

QY 574 LYEEARSAYLSKELASLLRAGGVLAYELLPALDEVLASDRFLGLGWLEQ--ARAAVS 631
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 L-LEEKRSALLRAE-----ELETALMEIVKEIDNRRLQLSAKVQEQEAEVQ 655

QY 632 EAEADFYEQ 640
   | : : : :
Db 656 RLSDKQEQ 664
```

Search completed: September 21, 2002, 12:24:47  
Job time: 62 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2002, 12:24:50 ; Search time 13.28 seconds  
(without alignments)  
2166.312 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAAGVLLLAGAGG.....VDLAKIFLKYPGWAGSV 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3939	100.0	743	1 ANAG_HUMAN	P54802 homo sapien
2	123.5	3.1	497	1 CPDH_MACFA	Q29488 macaca fasc
3	119	3.0	3649	1 ACVS_NOCIA	P27743 nocardia la
4	117.5	3.0	497	1 CPD6_HUMAN	P10635 homo sapien
5	110.5	2.8	1029	1 YK95_RHIME	Q52999 rhizobium m
6	109.5	2.8	1377	1 RHSA_ECOLI	P16916 escherichia
7	109.5	2.8	1411	1 RHSE_ECOLI	P16917 escherichia
8	108	2.7	726	1 CATA_SALTY	P17750 salmonella
9	104.5	2.7	726	1 CATA_ECOLI	P13029 escherichia
10	103	2.6	722	1 P85B_RAT	Q63788 rattus norv
11	101	2.6	698	1 YMCA_ECOLI	P75882 escherichia
12	101	2.6	3519	1 OL56_STRAT	Q07017 streptomyce
13	100	2.5	851	1 ENV_HV2D1	P17755 human immun
14	100	2.5	1323	1 NME4_MOUSE	Q03391 mus musculu
15	100	2.5	1336	1 NME4_HUMAN	O15399 homo sapien
16	99.5	2.5	1092	1 LIFR_MOUSE	P42703 mus musculu
17	99	2.5	624	1 SIR_SYNP7	P30008 synechococc
18	99	2.5	1230	1 UGSA_SOLTU	Q43846 solanum tub
19	99	2.5	1587	1 LMG3_HUMAN	Q9y6n6 homo sapien
20	98.5	2.5	512	1 NUOM_RHOCA	P50974 rhodobacter
21	98.5	2.5	1774	1 MSAS_PENPA	P22367 penicillium
22	98	2.5	860	1 ENV_HV2BE	P18094 human immun
23	98	2.5	1323	1 NME4_RAT	O62645 rattus norv
24	98	2.5	2067	1 B1MB_EMENI	P33144 emericella
25	97.5	2.5	497	1 GLPK_BACHD	Q9kdw8 bacillus ha
26	97.5	2.5	693	1 UL47_HSV1	P10231 herpes simp
27	97.5	2.5	786	1 AAC_ACTUT	P29958 actinoplane
28	97.5	2.5	842	1 VGLH_HSVBC	P27599 bovine herp
29	97.5	2.5	1078	1 C2TA_MOUSE	P79621 mus musculu
30	97	2.5	1094	1 EMBG_MYCTU	P72059 mycobacteri
31	97	2.5	2594	1 7LES_DROVI	P20806 drosophila
32	96.5	2.4	856	1 ENV_HV1MN	P05877 human immun
33	96	2.4	708	1 NICA_MOUSE	P57716 mus musculu

## ALIGNMENTS

RESULT 1

ID	ANAG_HUMAN	STANDARD	PRT	743 AA
AC	P54802;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Alpha-N-acetylglucosaminidase precursor (EC 3.2.1.50) (N-acetyl-alpha-glucosaminidase) (NAG).			
GN	NAGLU OR UFHSDI.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=96234097; PubMed=8650226;			
RA	Zhao H.G., Li H.H., Bach G., Schmidtchen A., Neufeld E.F.;			
RT	"The molecular basis of Sanfilippo syndrome type B.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:6101-6105(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.			
RX	MEDLINE=96372812; PubMed=8776591;			
RA	Weber B., Blanch L., Clements P.R., Scott H.S., Hopwood J.J.;			
RT	"Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis III B).";			
RL	Hum. Mol. Genet. 5:771-777(1996).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Zhao Z., Yazdani A., Shen Y., Sun Z.S., Bailey J., Caskey C.T.,			
RL	Lee C.C.;			
RN	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	VARIANTS MPS-IIIB H-92; S-115; C-140; K-153; L-358; V-664 AND R-682.			
RX	MEDLINE=98107938; PubMed=9443878;			
RA	Schmidtchen A., Greenberg D., Zhao H.G., Li H.H., Huang Y., Tieu P.,			
RA	Zhao H.-Z., Cheng S., Zhao Z., Whitley C.B., di Natale P.,			
RA	Neufeld E.F.;			
RT	"NAGLU mutations underlying Sanfilippo syndrome type B.";			
RL	Am. J. Hum. Genet. 62:64-69(1998).			
RN	[5]			
RP	VARIANTS MPS-IIIB.			
RX	MEDLINE=99133861; PubMed=9950362;			
RA	Bunge S., Knigge A., Steglich C., Kleijer W.J., van Diggelen O.P.,			
RA	Beck M., Gal A.;			
RT	"Mucopolysaccharidosis type IIIB (Sanfilippo B): identification of 18 novel alpha-N-acetylglucosaminidase gene mutations.";			
RL	J. Med. Genet. 36:28-31(1999).			
CC	!- FUNCTION: INVOLVED IN THE DEGRADATION OF HEPARAN SULFATE.			
CC	!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-acetyl-D-glucosamine residues in N-acetyl-alpha-D-glucosaminides.			
CC	!- SUBUNIT: MONOMER AND HOMODIMER.			
CC	!- TISSUE SPECIFICITY: LIVER, OVARY, PERIPHERAL BLOOD LEUKOCYTES,			
CC	!- TESTIS, PROSTATE, SPLEEN, COLON, LUNG, PLACENTA AND KIDNEY.			
CC	!- DISEASE: DEFECTS IN NAGLU ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS			

P23471 homo sapien  
P26279 erythroblast  
P14650 rattus norv  
P22089 burkholderi  
P08313 herpes simp  
Q10876 mycobacteri  
Q06457 klebsiella  
Q64403 cavia porce  
Q11001 manduca sex  
Q13796 homo sapien  
P08544 t genome po  
Q9y651 homo sapien

34 96 2.4 2314 1 PTP2\_HUMAN  
35 95.5 2.4 330 1 RCEN\_ERVSP  
36 95.5 2.4 914 1 PERT\_RAT  
37 95 2.4 344 1 LIC1\_BURCE  
38 94.5 2.4 664 1 UL47\_HSVIF  
39 94.5 2.4 761 1 CTPA\_MYCTU  
40 94.5 2.4 866 1 NASA\_KLEPN  
41 94 2.4 500 1 CPDG\_CAVPO  
42 94 2.4 990 1 AMPM\_MANSE  
43 94 2.4 1616 1 APXL\_HUMAN  
44 94 2.4 2303 1 POLG\_TMEVB  
45 93.5 2.4 276 1 SX21\_HUMAN

TYPE IIIB (MPS-IIIB) (ALSO KNOWN AS SANFILIPPO B SYNDROME), AN AUTOSOMAL RECESSIVE DISORDER WHOSE CLINICAL FEATURES ARE SEVERE MENTAL DETERIORATION BUT MILD SOMATIC MANIFESTATIONS IN CHILDHOOD, AND DEATH IN THE SECOND DECADE. BIOCHEMICALLY, THIS DISEASE IS CHARACTERIZED BY UNDEGRADED OR PARTIALLY DEGRADED HEPARAN SULFATE WHICH ACCUMULATES IN LYOSOMES AND IS EXCRETED IN URINE.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; U43572; AAC50512.1; -  
CC EMBL; U43573; AAC50513.1; -  
DR EMBL; U40846; AAB06188.1; -  
DR EMBL; L78464; AAB36604.1; -  
DR MIM; 252920; -  
KW Hydrolase; Glycosidase; Glycoprotein; Mucopolysaccharidosis;  
KW Signal; Disease mutation; Polymorphism.  
FT SIGNAL 1 23  
FT CHAIN 24 743  
FT CHAIN 59 743  
FT DOMAIN 68 71  
FT DOMAIN 84 87  
FT CARBOHYD 134 134  
FT CARBOHYD 261 261  
FT CARBOHYD 272 272  
FT CARBOHYD 435 435  
FT CARBOHYD 503 503  
FT CARBOHYD 526 526  
FT CARBOHYD 532 532  
FT CARBOHYD 79 79  
FT VARIANT 92 92  
FT VARIANT 100 100  
FT VARIANT 115 115  
FT VARIANT 140 140  
FT VARIANT 142 142  
FT VARIANT 153 153  
FT VARIANT 243 243  
FT VARIANT 277 277  
FT VARIANT 280 280  
FT VARIANT 292 292  
FT VARIANT 358 358  
FT VARIANT 452 452  
FT VARIANT 482 482  
FT VARIANT 561 561  
FT VARIANT 565 565  
FT VARIANT 643 643  
FT VARIANT 664 664  
FT VARIANT 674 674

ALPHA-N-ACETYLGLUCOSAMINIDASE 82 KDA  
FORM.  
ALPHA-N-ACETYLGLUCOSAMINIDASE 77 KDA  
FORM.  
POLY-GLY.  
POLY-ALA.  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
G -> C (IN MPS-IIIB).  
/FTID=VAR\_008979.  
Y -> H (IN MPS-IIIB).  
/FTID=VAR\_005007.  
H -> R.  
/FTID=VAR\_008980.  
P -> S (IN MPS-IIIB).  
/FTID=VAR\_005008.  
Y -> C (IN MPS-IIIB).  
/FTID=VAR\_005009.  
MISSING (IN MPS-IIIB).  
/FTID=VAR\_008981.  
E -> K (IN MPS-IIIB).  
/FTID=VAR\_005010.  
P -> L (IN MPS-IIIB).  
/FTID=VAR\_008982.  
C -> F (IN MPS-IIIB).  
/FTID=VAR\_008983.  
L -> P (IN MPS-IIIB).  
/FTID=VAR\_008984.  
G -> R (IN MPS-IIIB).  
/FTID=VAR\_008985.  
P -> L (IN MPS-IIIB).  
/FTID=VAR\_005011.  
E -> K (IN MPS-IIIB).  
/FTID=VAR\_008986.  
R -> W (IN MPS-IIIB).  
/FTID=VAR\_008987.  
L -> R (IN MPS-IIIB).  
/FTID=VAR\_008988.  
R -> Q (IN MPS-IIIB).  
/FTID=VAR\_008989.  
R -> H (IN MPS-IIIB).  
/FTID=VAR\_005012.  
A -> V (IN MPS-IIIB).  
/FTID=VAR\_005013.  
R -> H (IN MPS-IIIB).

FT VARIANT 682 682 /FTID=VAR\_005014.  
FT L -> R (IN MPS-IIIB).  
FT /FTID=VAR\_005015.  
FT E -> K (IN MPS-IIIB).  
FT /FTID=VAR\_008990.  
FT G -> R.  
FT /FTID=VAR\_008991.  
FT CONFLICT 551 551 A -> L (IN REF. 2; AA SEQUENCE).  
FT CONFLICT 553 553 S -> L (IN REF. 2; AA SEQUENCE).  
SQ SEQUENCE 743 AA; 82166 MW; 6D8D6A42C7BA6083 CRC64;  
  
Query Match 100.0%; Score 3939; DB 1; Length 743;  
Best Local Similarity 100.0%; Pred. No. 3.4e-279; Indels 0; Gaps 0;  
Matches 743; Conservative 0; Mismatches 0;  
  
QY 1 MEAVAVAAVGVLLLAGAGGAAGDEAREAAARVALVARLLGPGPAADFSVSVERALAAKP 60  
DB 1 MEAVAVAAVGVLLLAGAGGAAGDEAREAAARVALVARLLGPGPAADFSVSVERALAAKP 60  
QY 61 GLDYSYLGSGGAARVRVGRSTGVAAAAGLHRYLRDFCGCHVAVWSGSQLRPLRPVAPGGE 120  
DB 61 GLDYSYLGSGGAARVRVGRSTGVAAAAGLHRYLRDFCGCHVAVWSGSQLRPLRPVAPGGE 120  
QY 121 LTEATPNRYRYQNVCTOSYSFVWMDWARWEREDWMALNGINLALAWSGOEAIWQRYVL 180  
DB 121 LTEATPNRYRYQNVCTOSYSFVWMDWARWEREDWMALNGINLALAWSGOEAIWQRYVL 180  
QY 181 ALGLTQAEINEFFFTGPAFLAWGRMGNLHTWDGPPSWHIIKOLYLQHRVLDQMRSGMTP 240  
DB 181 ALGLTQAEINEFFFTGPAFLAWGRMGNLHTWDGPPSWHIIKOLYLQHRVLDQMRSGMTP 240  
QY 241 VLPFAGHVPEAVTRVFOVNVTKMSWGHENCSYSCSFLAPEDPIPIIGSLFLRELII 300  
DB 241 VLPFAGHVPEAVTRVFOVNVTKMSWGHENCSYSCSFLAPEDPIPIIGSLFLRELII 300  
QY 301 KEFGTDHYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTEAVLLQGMWLFQHQPF 360  
DB 301 KEFGTDHYGADTFNEMOPPSSEPSYLAATAATVAYEAMTAVDTEAVLLQGMWLFQHQPF 360  
QY 361 WGAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFOGQPIWCMHNFNGHGLFGAL 420  
DB 361 WGAQIRAVLGAVPRGRLLVLDLFAESQPVYTRTASFOGQPIWCMHNFNGHGLFGAL 420  
QY 421 EAVNGSGPEARLFPNSTVMGTGMAPEGISQNEVYSLMAELGWRKDPVPDLAAWYTSFAA 480  
DB 421 EAVNGSGPEARLFPNSTVMGTGMAPEGISQNEVYSLMAELGWRKDPVPDLAAWYTSFAA 480  
QY 481 RRYGVSHPDAGAAWRLRLRSVYNCSGEACRGNHRSPLVRRPSLQMNSTSIWYNRSDVFEAW 540  
DB 481 RRYGVSHPDAGAAWRLRLRSVYNCSGEACRGNHRSPLVRRPSLQMNSTSIWYNRSDVFEAW 540  
QY 541 RLLLTSAPLSATSPAFRYDLDDLTQAVQVELSVLYEEARSAYLSKELASLLRAGGVLAY 600  
DB 541 RLLLTSAPLSATSPAFRYDLDDLTQAVQVELSVLYEEARSAYLSKELASLLRAGGVLAY 600  
QY 601 ELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN 660  
DB 601 ELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFYEQNSRYQLTLWPGEGNILDYAN 660  
QY 661 KQLAGLVANYTTPRWKLFLEALVDSVAQGIPIQOQHFQDKNFTQLEQAFVLSKORYPSQPR 720  
DB 661 KQLAGLVANYTTPRWKLFLEALVDSVAQGIPIQOQHFQDKNFTQLEQAFVLSKORYPSQPR 720  
QY 721 GDTVDLAKKIFLKYYPGWAGSW 743  
DB 721 GDTVDLAKKIFLKYYPGWAGSW 743  
  
RESULT 2  
CPDH\_MACFA STANDARD; PRT; 497 AA.  
ID CPDH\_MACFA  
AC Q29488;

01-NOV-1997 (Rel. 35, Created)  
 01-NOV-1997 (Rel. 35, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 Cytochrome P450 2D17 (EC 1.14.14.1) (CYP2D17).  
 GN CYP2D17  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 CC Cercopithecinae; Macaca.  
 CC NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) -> ROH +  
 CC oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U38218; AAA79722.1; -;  
 CC HSP: P00179; 1D76.  
 CC InterPro: IPR001128; Cyt\_P450.  
 CC Pfam: PF00067; P450; 1.  
 CC PRINTS: PR00385; P450.  
 CC PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
 KW Microsome; Endoplasmic reticulum.  
 FT BINDING 443 443  
 FT HEME (BY SIMILARITY).  
 SQ SEQUENCE 497 AA; 56010 MW; 3594AA88F04E58B1 CRC64;

Query Match 3.1%; Score 123.5; DB 1; Length 497;  
 Best Local Similarity 22.9%; Pred. No. 0.16;  
 Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;

QY 343 TEAVLL-----OGWLFQHQPFQWPAQIRAVILGAVPRGRLLVLDLFAESQPVYTRT 394  
 Db 12 TVAIFLLVLDLMHRRQRWAARYPP---GPLPLPL-----GNLLHVD----- 50  
 QY 395 ASFGQGFPTWCMLEHNFSGN-----HGLFGALEA-VNGGPPAARL--FPNST 437  
 Db 51 --FKNTPYCFDQLRRFGNVFSLQLAWTPVVVINGLAARVREALVTCGEDTADRPVPIQ 108  
 QY 438 MVGTGMAPEGI-----SQNEVYVSLMAELGWRKDPVDPDLAAWVTSFAARYGVSH 487  
 Db 109 VLGFGRSOGVFLARYGPAPWRQRFSVSTLNGLGK--KSLEQWTEEAACLCAPT 165  
 QY 488 PDAGAAWR---LLRSVYVNCSEACRGHNSPLVRRPSLQMWTSIWNRSVDVFEAWRL 544  
 Db 166 DQGRFRFNSLLDKAVSN-----VIASLTGR----- 193  
 QY 545 TSAPSLATSPAFRYD-----LLDLTROAVQELVSLVYEEARSAYLSKELAS---LLR-- 593  
 Db 194 -----REYDDPRLRLFDLTHEALKE-----ESGLFRELVLNAIPLLLRIP 234  
 QY 594 --AGGVLAYE--LLPALDEVL-----ASDSRFLGLSWLEQARAAAVSEAEADFYEQ 640  
 Db 235 GLAGKVLRSQKAFATQLODELLTHERWTPDPAQPPRDLTEAFLEAEKAK-GNPSSFNEE 293  
 QY 641 NSRYQTLTWGPBGNILDVANKQAGLVANYTPRRRLFLEALVDVSAQIPFQHQFDKN 700  
 Db 294 NLR-----WVADLFSAGMVTSTTLAWGLLMLILHDPVQRRV---QOEIDV 338

QY 701 VFOLEQAFVLSKQRYP-----SQPRGDTVDL 726  
 Db 339 IGQVRPEMGDQARMPYTTAVIHEVQRFQDIVPL 372  
 RESULT 3  
 ACVS\_NOCLA  
 ID ACVS\_NOCLA STANDARD; PRT; 3649 AA.  
 AC P27743;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Delta-(L-alpha-aminoadipyl)-L-cysteine-D-valine synthetase  
 DE (EC 6.-.-.-) (ACV synthetase) (ACVS).  
 GN PCBAB.  
 OS Nocardia lactamdurans.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;  
 CC Anycolatopsis.  
 CC NCBI\_TaxID=1913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VAR LC 411;  
 RX MEDLINE=92065808; PubMed=1956290;  
 RA Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;  
 RT "The cephamycin biosynthetic genes pcbAB, encoding a large  
 RT multidomain peptide synthetase, and pcbC of Nocardia lactamdurans are  
 RT clustered together in an organization different from the same genes  
 RT in Acromonium chrysogenum and Penicillium chrysogenum.";  
 RT Mol. Microbiol. 5:1125-1133(1991).  
 RL -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE  
 CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS  
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER  
 CC INTERMEDIATES.  
 CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND  
 CC CEPHALOSPORIN.  
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
 CC FAMILY.  
 CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X57310; CAA40561.1; -;  
 CC PIR: S18268; S18268.  
 CC HSP: P14687; 1AMU.  
 CC InterPro: IPR000873; AMP-bind.  
 CC InterPro: IPR001242; DUF4.  
 CC InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
 CC InterPro: IPR003880; Phosphoant\_attach.  
 CC InterPro: IPR001031; Thioesterase.  
 CC Pfam: PF00501; AMP-binding; 3.  
 CC Pfam: PF00668; Condensation; 3.  
 CC Pfam: PF00550; pp-binding; 3.  
 CC Pfam: PF00975; Thioesterase; 1.  
 CC PRINTS: PR00154; AMPBINDING.  
 CC PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.  
 CC PROSITE: PS00455; AMP\_BINDING; 1.  
 CC PROSITE: PS00075; ACP\_DOMAIN; 3.  
 CC Ligase; Antibiotic biosynthesis; Multifunctional enzyme;  
 CC Repeat; Phosphopantetheine.  
 CC REPEAT 401 861  
 CC REPEAT 1014 1937  
 CC REPEAT 2079 2985  
 CC DOMAIN 788 857  
 CC DOMAIN 1864 1933  
 CC DOMAIN 2910 2981  
 CC DOMAIN 1 (ADIPATE-ACTIVATING).  
 CC DOMAIN 2 (CYSTEINE-ACTIVATING).  
 CC DOMAIN 3 (VALINE-ACTIVATING).  
 CC ACYL CARRIER (ACP) 1.  
 CC ACYL CARRIER (ACP) 2.  
 CC ACYL CARRIER (ACP) 3.

FT	BINDING	820	820	PHOSPHOPANTETHEINE (BY SIMILARITY).
FT	BINDING	1896	1896	PHOSPHOPANTETHEINE (BY SIMILARITY).
FT	BINDING	2944	2944	PHOSPHOPANTETHEINE (BY SIMILARITY).
FT	ACT_SITE	3502	3502	THIOESTERASE (BY SIMILARITY).
SEQ	SEQUENCE	3649 AA; 404079 MW; 6FD095704F858EB6 CRC64;		
Query Match 3.0%; Score 119; DB 1; Length 3649;				
Best Local Similarity 21.1%; Pred. No. 4.9;				
Matches 174; Conservative 63; Mismatches 283; Indels 306; Gaps 37;				
QY	16	AGAGGAACDEAREAAARALV-----ARLGGCPAADEFVSVERALAAKPGDLYSLGG	69	
DB	1988	AALPGALGTLVRRHPALRTLTKTDGQVRRYPPIADDVRLVLP-----	2031	
QY	70	GGAARVVRGSGTGVAAAAGLHRLYRDFCGCHVAVSGSQLRPLRPAPVPGELTEATPNRY	129	
DB	2032	-----STTVDSRAELDEVLTE-----RAGVYFRLHLELP-----RAEAFDHGD	2070	
QY	130	RYQNYCTQSYFVWHDWARHREI-----DMMALINGINALAWSQGAIAWQVY	179	
DB	2071	EIYLSVVVHHSCFDGMSWDFIREALAAALDGVPEADGALRG-----YGFAYWQRY	2124	
QY	180	LALGLTQAEINEFFTGAFLAIRMGNLHT-----WDGPLPESWHIK-----QLYLQHRVLD	231	
DB	2125	LT-GKRLAALTEWT-----GALGGFETIALPLDHPPRPREDYRGREFELFELDERTT	2176	
QY	232	QMRSGFWT-----	239	
DB	2177	ALRELARTARVSLYSVLLGAWCLMLNMYTGQHDLVVGTGVSANRGPRFEDRAVGFFANLLA	2236	
QY	240	-----PVLPAFAGHPVEATRVFPQVNTTKMSGWHFNCYSYSCSFLLAPEDPI-FPI	290	
DB	2337	LRVRDPAATLPAYRSVGEAV--VAAQVH-----GELPFEQLVRELKVEDPSRHP	2287	
QY	291	IGSLFRELIKERGTOHHYGAOTFENMOPSPSESYLAAATVAYEAMTAVDTAEVWLLQ	350	
DB	2288	LQLNFTLQNV-----SDH-----TSALTGYQDPS-----GGWTTTKFD-LSAATMTATGLA	2333	
QY	351	GWL-----FQHQPFQWGPQAQIRAVLGAVPRGRLLVLDFLAFESQ-PVY	391	
DB	2334	GNLYTAAASLFDDTSAGSFTATFKHV-----LAEFASAAQTFIAQLTALDEPGAALPDA	2388	
QY	392	TRTASFQGPFIWCLMHNFGNHLFGALEAVNGGPEAARLPNPSWVG-----MAPEGI	448	
DB	2389	TRRARPPGPP-----GRCTRLE-----EEVAATWPDVAVVHGDVRLTYREL	2430	
QY	449	SQ--NEVVYSLMAELGHRKDP-----VPDLAAAVTSPAARYGVSHPDGAA	493	
DB	2431	NERANRLAHLHRSVAEPRADIELIALVDKSELTVALIAVWAKAGAAVYMPIDSPIDDR	2490	
QY	494	W-----RLLLRSVYNCSGEACGRHNRSPLVRRPFSQNMNTSIWYNRSDVFPEARLLITS	546	
DB	2491	FMLSDTGAKLVL-----AGEAHGSRVG-----LTSGDVLDLEQLDLTG	2529	
QY	547	APSLATSPAFRYDLDLTROAVOELVSLYYEARSAYLSKELASLLRAGGVLAYELLPAL	606	
DB	2530	EP--AENP-----VTETSTELAYAIY---TSGTCKPKAVILVSHGVS-----	2567	
QY	607	DEVLASDSRFLGWSLQAAAAVSAEADDFEONSRQYLTLMWGPENIL-----	656	
DB	2568	DSFRAQLSGRYFGSPDESAAVLFL---ANYVDFDSVEQALSLVGGHKLVLPPPSAADD	2624	
QY	657	----DYANKQIAGLVANYTTPRWRLFLEALVDSVAQIFQHQHFD	698	
DB	2625	PAFYELANRE--GL-----SYLSGTPQVVERFD	2650	

RESULT	4				
CPD6_HUMAN					
ID	CPD6_HUMAN	STANDARD;	PRT;	497 AA.	
AC	P10635; Q16752;				
DT	01-JUL-1989 (Rel. 11, Created)				
RP	VARIANT ILE-10/ (CYP2D6*17).				
RX	MEDLINE=97126511; PubMed=8971426;				
RA	Masimirembwa C., Persson I., Bertilsson L., Hasler J.,				
RA	Ingelman-Sundberg M.;				
RT	"A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a				







comparison of rhsA with other members of the rhs multigene family.";  
[2]  
J. Bacteriol. 172:446-456(1990).  
SEQUENCE FROM N.A.  
STRAIN=K12 / MG1655;  
MEDLINE=94316500; PubMed=8041620;  
Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
"Analysis of the Escherichia coli genome. V. DNA sequence of the  
region from 76.0 to 81.5 minutes.";  
Nucleic Acids Res. 22:2576-2586(1994).  
[3]  
REVIEW  
MEDLINE=95020608; PubMed=7934896;  
Hill C.W., Sandt C.H., Vlazny D.A.;  
"Rhs elements of Escherichia coli: a family of genetic composites  
each encoding a large mosaic protein.";  
Mol. Microbiol. 12:865-871(1994).  
CC -!- FUNCTION: RHS ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
CC -!- DOMAIN: EACH RHS APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa  
AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
CC -!- SIMILARITY: BELONGS TO THE RHS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L19044; AAC95065.1; -;  
DR EMBL: U00039; AAB18570.1; -;  
DR EMBL: AE000437; AAC76617.1; -;  
DR EcoGene: EG10846; rhsA.  
DR InterPro: IPR001826; RHS.  
DR PRINTS: PR00394; RHSPROTEIN.  
KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
FT SIGNAL 1 26  
FT CHAIN 27 1377  
FT TRANSMEM 28 55  
FT DOMAIN 330 1186  
FT REPEAT 330 352  
FT REPEAT 353 374  
FT REPEAT 375 417  
FT REPEAT 418 438  
FT REPEAT 439 460  
FT REPEAT 461 481  
FT REPEAT 482 502  
FT REPEAT 503 525  
FT REPEAT 526 546  
FT REPEAT 547 567  
FT REPEAT 568 588  
FT REPEAT 589 609  
FT REPEAT 610 629  
FT REPEAT 630 650  
FT REPEAT 651 671  
FT REPEAT 672 691  
FT REPEAT 692 711  
FT REPEAT 712 734  
FT REPEAT 735 758  
FT REPEAT 759 788  
FT REPEAT 808 828  
FT REPEAT 829 850  
FT REPEAT 851 871  
FT REPEAT 872 894  
FT REPEAT 895 930  
FT REPEAT 931 959  
FT REPEAT 960 984  
FT REPEAT 985 1019  
FT REPEAT 1162 1186  
SQ SEQUENCE 1377 AA; 156320 MW; 21ACA989E74200FE CRC64;

Query Match 2.8%; Score 109.5; DB 1; Length 1377;  
Best Local Similarity 20.8%; Pred. No. 6.5;  
Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
QY 57 AAKPGDITYSLGGG--GAARVRVRSSTGVAAA---AGL--HRYLRDFCGCHVAVSGSQL 108  
DB 6 AARQG-DMTQYGGISVQGSAGVRICAGTGVACSGPGGVTSHPVNPILLGAKVLPGETDI 64  
QY 109 RLPRPLPAV-----PG-----ELTEAT-----PNRYRYQ 133  
DB 65 ALPGPLPILSRTSYSTRTKTPAPVGLSGPWKMPADIRLQLRONTLILSDNGGSLTFE 124  
QY 134 NVC--TQSYSPVWWDWAREREIDWALNGINLAWSGQEARIVRV-----YLA 181  
DB 125 HLFPGEDGYS-----RSESLWLVGGVAKLDEGHRLAALWQALPEELRLSPHYLA 175  
QY 182 LGLTQAEINEFTGPAFLAWGRNLTW-----DGPLPPSHHIQOLYLRHVL 230  
DB 176 TNSPQ-----GP-----WLLG-----WCERVPEADVLPALEP-----YRL 209  
QY 231 DOM-RSFGMTPLVPFAFGHVPEAVTRVPQVNTKMSGWHFNCYSYCSFLLAPEDPIFP 289  
DB 210 TGLVDRFRGTQTFHR-----EAAGEFSGEITGVTGAWRHFRLVLTTOAQAEE----- 258  
QY 290 IIGSLFLRELKKEGTDIHYGADTFNEMQPSSEPSYLAATTAAYEAMTAVDTAVWLL 349  
DB 259 -----ARQQAISGTE-----PSAFPTLPGYTE--YGRDNGIRLSAVWL- 296  
QY 350 QGWLFPQHPQFQWPAQIRAV-----LGAVPRGRLLVLDLFAESQPVYTRTASFQGPFTWC 405  
DB 297 -----THDPEY--PENLPAAPLVRYGWTGPRGELAV-----YDRSGK----- 331  
QY 406 MLHNEGGNHGLFGALEA--VNGGPEAAARLFNSTWVGTMGAPEGIS----- 449  
DB 332 QVRSFTYDDKYRGRVVAHRHTGRPEIRYRYSDSGRVTEQLNPAGLSYTYQEKDRITTD 391  
QY 450 ---QNEVYISLMAELGWRK-----DVPDLAAWVTSFAARRYGVSHPD 490  
DB 392 SLDRREVLHT--QGEAGLKRVRVKEHADGSYTSQSFDAVGLRLA--QTDAGRTEYSPDV 448  
QY 491 GAANRLLRLSYVNCSEAC-----RGHNRSLVRRP-SLQWNTSIWNRSDVFEARLLLT 545  
DB 449 VTG---LITRITTPDGRASAFYNNHNLTSATGPDGLEL-----RREYDELGRLIQE 498  
QY 546 SAPSLATSPAPRYD 559  
DB 499 TAPD-GDITRYRD 511  
RESULT 7  
RHSB\_ECOLI  
ID RHSB\_ECOLI STANDARD; PRT; 1411 AA.  
AC P16917; P76701;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE RhsB protein precursor.  
GN RHSB OR B3482.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=93259920; PubMed=8387990;  
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
"Rhs elements of Escherichia coli K-12: complex composites of shared  
and unique components that have different evolutionary histories.";  
J. Bacteriol. 175:2799-2808(1993).  
RN [2]  
RP REVISION TO 405.  
RA Hill C.W.;

Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

RL [3] SEQUENCE FROM N.A.  
RN STRAIN-K12 / MG1655;  
RC MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [4] SEQUENCE FROM N.A.  
RP STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [5] SEQUENCE OF 1-100 FROM N.A.  
RP STRAIN-K12;  
RX MEDLINE-89123133; PubMed-2644231;  
RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
RT "rbs gene family of *Escherichia coli* K-12.";  
RL J. Bacteriol. 171:636-642(1989).  
RN [6] SEQUENCE OF 1221-1411 FROM N.A.  
RP STRAIN-K12;  
RX MEDLINE-90094253; PubMed-2403547;  
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
RT "Structure of the rbsA locus from *Escherichia coli* K-12 and  
comparison of rbsA with other members of the rbs multigene family.";  
RL J. Bacteriol. 172:446-456(1990).  
RN [7] REVIEW.  
RP MEDLINE-95020608; PubMed-7934896;  
RX Hill C.W., Sandt C.H., Vlazny D.A.;  
RT "Rbs elements of *Escherichia coli*: a family of genetic composites  
each encoding a large mosaic protein.";  
RL Mol. Microbiol. 12:865-871(1994).  
CC -1- FUNCTION: RBS ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
CC -1- DOMAIN: EACH RBS APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa  
AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
CC -1- SIMILARITY: BELONGS TO THE RBS FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L02370; AAC61883.1; -;  
CC EMBL; U00039; AAB18457.1; -;  
CC EMBL; AE000424; AAC76507.1; -;  
CC EcoGene; EG10847; rbsB.  
CC InterPro; IPR001826; RBS.  
CC PRINTS; PR00394; RBSPROTEIN.  
CC Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
KW SIGNAL 1 26  
FT CHAIN 27 1411  
FT RBSB PROTEIN.  
FT POTENTIAL.  
FT DOMAIN 330 1186  
FT TRANSMEM 28 55  
FT REPEAT 330 352  
FT REPEAT 353 374  
FT REPEAT 375 417  
FT REPEAT 418 438  
FT REPEAT 439 460  
FT REPEAT 461 481

FT REPEAT 482 502 7.  
FT REPEAT 503 525 8.  
FT REPEAT 526 546 9.  
FT REPEAT 547 567 10.  
FT REPEAT 568 588 11.  
FT REPEAT 589 609 12.  
FT REPEAT 610 629 13.  
FT REPEAT 630 650 14.  
FT REPEAT 651 671 15.  
FT REPEAT 672 691 16.  
FT REPEAT 692 711 17.  
FT REPEAT 712 734 18.  
FT REPEAT 735 758 19.  
FT REPEAT 808 828 20.  
FT REPEAT 829 850 21.  
FT REPEAT 851 871 22.  
FT REPEAT 872 894 23.  
FT REPEAT 895 930 24.  
FT REPEAT 931 959 25.  
FT REPEAT 960 984 26.  
FT REPEAT 985 1019 27.  
FT REPEAT 1162 1186 28.  
FT CONFLICT 1130 1130 Q -> K (IN REF. 4).  
SQ SEQUENCE 1411 AA; 159394 MW; 567AC4EE713D9E07 CRC64;  
  
Query Match 2.8%; Score 109.5; DB 1; Length 1411;  
Best Local Similarity 20.8%; Pred. No. 6.7;  
Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
  
QY 57 AAKPGLDIYSLGGG---GAARVRVGRSGTGVA---AGL---HRYLRDFCGCHVAMSGSQL 108  
DB 6 AARQG-DMTQYGGSIQVSAGVRIGAPTGVACSVCPGGVTSGHPVNPILGAKVLPGETDI 64  
QY 109 RLPRPLPAV-----PG-----ELTEAT-----PNRYRYQ 133  
DB 65 ALPGPLPFLISRTYSYRTKTPAPVGSGLPGWKMPADIRLQLRONTLILSDNGGRSLYFE 124  
QY 134 NVC---TQSYFVWDWAKWERIDWALNGINLALWGOEAIQRV-----YLA 181  
DB 125 LFLGPDGDSY-----RSESLWLRGVGVAKLDEGHRLAALWQALPEELKLSPHRYLA 175  
QY 182 LGLTAAEINEFTTGAFLAWGMNLTW-----DGPLPPSWHIKOLYLQHRVL 230  
DB 176 TNSPQ-----GP-----WLLG---WCEVRPEADVLPAPLPP-----YRVL 209  
QY 231 DQM-RSFGMTPLVPAFAGHVPVAVTRVPQVNVNMGSHGHCNCSYCSFLAPEDPIFP 289  
DB 210 TGLVDRFGRQTQFHR-----EAAGEFSGEITGVTDGAWRHFRLLVTTQAQRAEE--- 258  
QY 290 IIGSLFLRELKEFGDHIYGADTFNEMOPSPSEPSYLAATAVVEAMTAVDTAVWLL 349  
DB 259 -----ARQAISGGTE-----PSAFDPTLPGYTE--YGRDNGIRLSAVWL- 296  
QY 350 QGMFLQFOQFMGPAQIRAV---LCAVPRGRLLVLDLFAESQPVYVTRTASFQGGPFIMC 405  
DB 297 -----THDPEY--PENLPAALVRYGWTGRELAV-----VYDRSGK----- 331  
QY 406 MLHNFNGNHLFGALEA---VNGGPEAARLPFNSTWVGTGMAPEGIS----- 449  
DB 332 QVRSFTYDDKYRGMVVAHRHTGRPEIRYRYSDDGRVTEQLNPNAGLSYTYQYKDRITITD 391  
QY 450 ---QNEVWYSLMAELGWRK-----DPVPDLAAVWTSFAAIRYGVSHPDA 490  
DB 392 SLDRREVLHT-QGEAGLKRVRKKEHADGVSQTQSFDAVGLRA--QTDAAGRTTEYSPOV 448  
QY 491 GAARLLLRVSVNCSEAC-----RGNHRSPVLRPP-SLOMNTSIWYNRSDVFEAMRLILT 545  
DB 449 VTG---LITRITTPDGRSAFYNNHNLQTSATGPDGLEL-----RREYDELGLIOE 498  
QY 546 SAPSLATSPAFRYD 559  
DB 499 TAPD-GDITRYRYD 511

```
RESULT 8
ID CATA_SALTY STANDARD; PRT; 726 AA.
AC P17750;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
GN KATG OR STM4106.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91117169; PubMed=2277629;
RA Loewen P.C., Stauffer G.V.;
RT "Nucleotide sequence of katG of Salmonella typhimurium LT2 and
characterization of its product, hydroperoxidase I.";
RL Mol. Gen. Genet. 224:147-151(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RT Nature 413:852-856(2001).
CC 1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
BROAD-SPECTRUM PEROXIDASE ACTIVITIES.
CC 1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC 1- COFACTOR: TWO PROTOHEME IX GROUPS PER TETRAMER.
CC 1- SUBUNIT: HOMOTETRAMER.
CC 1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
PEROXIDASE/CATALASE SUBFAMILY.
CC 1- SIMILARITY: CONTAINS 1 HEME REGULATORY MOTIF (HRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X53001; CAA37187.1; -.
CC EMBL; AE008891; AAL22946.1; -.
CC PIR; S12039; CSEBHT.
CC HSP; P48534; IAPX.
CC StyGene; SG10190; katG.
CC InterPro; IPR002016; Peroxidase.
CC Pfam; PF00141; peroxidase; 1.
CC PROSITE; PR00458; PEROXIDASE.
CC PROSITE; PS00435; PEROXIDASE_1; 1.
CC PROSITE; PS00436; PEROXIDASE_2; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
complete proteome.
FT ACT_SITE 102 BY SIMILARITY.
FT ACT_SITE 106 106
FT BINDING 267 PROXIMAL HEME LIGAND (BY SIMILARITY).
FT DOMAIN 15 20 HRM 1 (POTENTIAL).
FT CONFLICT 71 71 Y -> YY (IN REF. 1).
FT CONFLICT 136 136 N -> T (IN REF. 1).
FT CONFLICT 223 223 G -> D (IN REF. 1).
FT CONFLICT 228 228 N -> T (IN REF. 1).
```

```
FT CONFLICT 233 233 D -> N (IN REF. 1).
FT CONFLICT 275 275 A -> P (IN REF. 1).
FT CONFLICT 545 545 G -> R (IN REF. 1).
FT CONFLICT 549 549 S -> H (IN REF. 1).
FT CONFLICT 553 553 A -> P (IN REF. 1).
FT CONFLICT 561 561 Q -> H (IN REF. 1).
FT CONFLICT 634 634 R -> K (IN REF. 1).
SQ SEQUENCE 726 AA; 79656 MW; 7C4BA44439E9FFAB CRC64;

Query Match 2.7%; Score 108; DB 1; Length 726;
Best Local Similarity 17.7%; Pred. No. 3.5;
Matches 134; Conservative 73; Mismatches 196; Indels 352; Gaps 33;

QY 86 AAGLHRYLDFCGCHVAMSGSOLRLPRLPAPVPGBELTEATPNR-----YRYQNVCTQ 138
DB 26 SAGAGTASRD-----WPNQLRV-----DLLNQHSNRSNPLGDEDFYRKEFSKL 69
QY 139 SYSFV-----WWDWARWEREIDMAMNGINLALAWS----- 169
DB 70 DYSALKGDLKALLTDSQPWW-----PADMGSYVGLFIRMAWHGAGTYRSIDGRGGAGR 122
QY 170 -----GQEAIWQVRVYLALGLTOAEINEFFTGPAF 198
DB 123 GQORFAPLNSWPDNVSLDKARLLWPIKQYQKISWADLFLAGNVALENSGFT---- 178
QY 199 LAWGRMGNLHWGDLPPSPWHIKQYLQHRVLDQMRSEGMTFVLPAPAFAGHVPEAVTRYFP 258
DB 179 FGFG-AGREDWEPDLVDNWGDEKAWLTHR-----H-PEALAKA-- 215
QY 259 QVNTKMGSWGHNCSYCSFLLAPEDPIFPIGLSLFLRELKKEGTHDIYGAFTFNEMQ 318
DB 216 PLGATEMG-----LIYVNPGE-----DH----- 234
QY 319 PPSSEPSYLAATTAVYEAMTAVTEAVWLQGLWFQHPQFQWGPAPQRAVLGAVPRGRL 378
DB 235 --SGEPLSAAAIATFGNMGMDNETVALLAG---GHTLGTKHCAAAASHVGADP--- 285
QY 379 LVLDLFAESQPVYRTASFQOPFWCMLHNPFGNGHGLFGALEAVNGGPEAA----- 430
DB 286 -----EAAPI-----EAQGLGWASSYSGS---VGA-DATISGLEVVVWTQTPTQW 325
QY 431 -----RLFPNSTVMGTGMAPEGISONEVYS-----LMAELGWRKD 466
DB 376 SNYPFENLFKYEYV--QTRSPAGAIQEAVDAPDILIPDPDPKRRKPTMLVTLDTLRFD 383
QY 467 P-----VDDLAAWVTSFAARRYGVSHPDAGAAWRLLRSLVYNCSGEACRGNHRSPL 517
DB 384 PEFKISRRLNDPQAFNEAFARAWFKLTHRDMGPKARYI----- 423
QY 518 VRRPSLOWNTSIWYNSRSDVFEAWRLLLTSAPSLATSPAFRYDLDL-----TRQAVOEL 571
DB 424 --GPEVPKEDLIWOD-----PLPQPLYQPT-QEDIINLKAATAASGLISEM 467
QY 572 VSLYYEEARSAYLSKELASLIRAG-----GVLAYELLPALDEVL 610
DB 468 VSV-----AWAS---ASTFRGDKRGKANGARLALAPQBDVDVNAVARVLEIGI 517
QY 611 ASDSRFLGS-----WLEQARAAA-----VSEAEAD 636
DB 518 KTTNKASLADIIVLAGVVGIEQAAAAAGVSVTFAPGRVDARQDQTDIEMFSLLEPIAD 577
QY 637 FYEQ-----NSRYQLTILWGPEGNII 656
DB 578 GFRNYRARDVSTTESLLIDKAOQLTLTAPEMTVL 612

RESULT 9
CATA_ECOLI
ID CATA_ECOLI STANDARD; PRT; 726 AA.
AC P13029;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
```





```
QY 108 LRLPRLPVAVGE--LTATPNRRYYQ-----NVCTQSYFVWMDWA- 148
D 108 LRLPRLPVAVGE--LTATPNRRYYQ-----NVCTQSYFVWMDWA- 148
D 43 LOMPARIAGEFSVNRDNDQYREYTSVALPFWLECTIRYDTRKYS-QWEDFS 101
QY 149 -----RWEREIDWMLNGINLALAWSGQAIQWVYVLAJGL-----T 185
D 102 DQSYKDKSDFKRLWEGY-----WLPQVAFGRDIAGTGLDFGEVLYASK 148
QY 186 QAEINEFFTPAFLAWGRGNLHTWDGPLPSPWHIKYLYLQHRV-----LDQMS 235
D 149 QAGPEFTLG---MANGVAGNAGNITNPF---CRVSDKYC-HRAESHDAGDISFSDIFRG 201
QY 236 ---FG---WTPVLP-----PAGHVPEAVTRVPQVNVVTKMGSGHENC 273
D 202 PASIFEGIEYQTPWPLRLKLEYDGNVQNDPAGKLPQ---SHFNVGAVYRAASWADNL 259
QY 274 SYSCSFLAPEDPIPIIGSLFLRELKEFGTDHIYG---ADTFNEMQP---PSSEPSYL 327
D 260 SY-----ERGNLTFMFGTTLRTNFENLDRPALRDTKPAYQ 293
QY 328 AAT-----TAVYEMTAV-----DTEAVVLQWLFQHPQFQWGAQIRAVLGAVPR 375
D 294 PAPESEGLQYTVANOLTKYKYNAGDAPEIQLODKTLYMSGOQY---KYRDSREAVDR 349
QY 376 -GRLLVLDLFAESQPVYTRTASFGQOPFIWCMHNFNGNHLFGALEAVNGGPEAARLFP 434
D 350 ANRILVNNL-----P-----QVKEI-----SVTKREHMAVMT 378
QY 435 NSTMVCT-----GMAPEISQNEVYSLMAELGNRKDPVPLAAMVTSFAARRGVSH 488
D 379 TETDVASLRKQLAGTAP---GOSEPLQOORVE-----AEDLSAFGRGRIREDRFSYS 428
QY 489 DAGAANRLLSRYNCSGECRHNRSPLVRPSLQMNITSW-----YNR 533
D 429-----FNPTLSQSGPEDFTNFQGLMSSARYWFTDHLDDGGIFTNIYN 475
QY 534 SDVFEAWRLILTSAPSLSATSPAPRYDLDDLTQAV-----OELVSLYEEARSAY 583
D 476 YDFKSL-----PADSTLPVRYTHIRDVVRNDVYLNQANFYADLGNFGVQYGGY 530
QY 584 LSKELASLRAGVLAPELLPALDEVLASDRSLGSLWLEQARAAVSEAEADFFQNSR 643
D 531 LETMYAGV---GSELLYRPL-----DACWALG-----VDVNYVKORD- 564
QY 644 YQLTLMGPEGNILDYANKQLAGLVANYTTPRWRLFLLEALVDSVAQGIFFQOHQDNVQ 703
D 565 -----WDMMRFTDYSTP---TGFVYAWNP-----PTLNGVLMK----- 596
QY 704 LEQAFVLSKORYPSQPRGDTVDLAKKIFLYKYPGWAGSW 743
D 597 -----LSVGOYLAKDKGATIDVAK-----RFDSCGAVGVW 626
```

```
RESULT 12
OL56_STRAT STANDARD; PRT; 3519 AA.
ID OL56_STRAT
AC Q07017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oleandomycin polyketide synthase, modules 5 and 6.
GN ORF6.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150470; PubMed=8107683;
RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT "Characterisation of a Streptomyces antibioticus gene encoding a type
I polyketide synthase which has an unusual coding sequence.";
```

```
RL Mol. Gen. Genet. 242:358-362(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC LACTONE RING.
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L09654; AAA19695.1; -.
CC HSP; P25715; IMLA.
CC InterPro; IPR001227; Acyltransf_domain.
CC InterPro; IPR000794; Ketoacyl-synt.
CC InterPro; IPR003880; Phosphopant_attach.
CC InterPro; IPR001031; Thioesterase.
CC Pfam; PF00698; Acyl_transf; 2.
CC Pfam; PF02801; ketoacyl-synt_C; 2.
CC Pfam; PF00550; pp-binding; 2.
CC Pfam; PF00975; Thioesterase; 1.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
CC PROSITE; PS00606; B_KETOACYL-SYNTHASE; 2.
CC PROSITE; PS50075; ACP_DOMAIN; 2.
CC Transfaser; Acyltransferase; Antibiotic biosynthesis; NADP;
CC Phosphopantetheine; Multifunctional enzyme; Repeat.
CC DOMAIN 1 ? MODULE 5.
CC FT DOMAIN 1 ? 3519 MODULE 6.
CC FT DOMAIN 32 501 BETA-KETOACYL SYNTHASE 1.
CC FT DOMAIN 569 890 ACYLTRANSFERASE (AT) 1.
CC FT DOMAIN 1200 1382 BETA-KETOACYL REDUCTASE 1.
CC FT DOMAIN 1487 1561 ACYL CARRIER (ACP) 1.
CC FT DOMAIN 1686 2156 BETA-KETOACYL SYNTHASE 2.
CC FT DOMAIN 2220 2541 ACYLTRANSFERASE (AT) 2.
CC FT DOMAIN 2856 3038 BETA-KETOACYL REDUCTASE 2.
CC FT DOMAIN 3141 3215 ACYL CARRIER (ACP) 2.
CC FT ACT_SITE ? ? ACYL-ENZYME INTERMEDIATE.
CC FT BINDING ? ? PHOSPHOPANTETHEINE (BY SIMILARITY).
CC FT DOMAIN 3270 3519 THIOESTERASE.
CC FT ACT_SITE 210 210 BETA-KETOACYL SYNTHASE.
CC FT ACT_SITE 660 660 ACYL-ENZYME INTERMEDIATE.
CC FT NP_BIND 1203 1249 NADP.
CC FT BINDING 1524 1524 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC FT ACT_SITE 1859 1859 BETA-KETOACYL SYNTHASE.
CC FT ACT_SITE 2311 2311 ACYL-ENZYME INTERMEDIATE.
CC FT NP_BIND 2859 2905 NADP.
CC FT BINDING 3178 3178 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC SEQUENCE 3519 AA; 368561 MW; 41AE78AAAE61F86 CRC64;
```

```
Query Match 2.6%; Score 101; DB 1; Length 3519;
Best Local Similarity 19.3%; Pred. No. 95;
Matches 122; Conservativity 63; Mismatches 234; Indels 214; Gaps 28;
```

```
QY 4 VAVAAVGVLLLAGAGGAAGDAREAAVRLVARLLGPGPAADFSVSVERALAAKPGLD 63
D 664 IAAACVAGGSLA-----DGAARVVLRSAIRATAGGGWVSVLPAGRVRTMLEED 716
QY 64 TYSLGGGAARVRVGRSTGVAAAAGLHRYLRDFCGCH-----VAMSGSQLRLPR 113
D 717 -----GRISVAVNGPSSVTSVGVQVQALDELLAGCEREGVRARVPVDYASHSQMDQL 770
QY 114 LPAPVGELTEATPNRYRYQNVCTQSYFVWMDW-----ARWERED 155
D 771 RDELLEALADITP-----QDSSVPFFSTVTADMLGTALGAGYWFNTLRVRFQEA 824
QY 156 WMAINGINLALAWSGQ-----EALWQR--VYALGLUTQAEINEFFTGPAFLANG 202
D 825 GLVQAQMGAFVCECPHPLVPGIEQTLQDLDALDQNAVFGSLRDEGGDLRFLTS----- 877
```



[illegible]

```
QY 642 SRYQLTACGP-----EGNILDYANKQLAGL-----VANYTPTWR--- 676
DB 553 EMLRLTVGKTKNQLARVTAIEKYLKDOQLNSWCAFRVCHTTPVWVNDLSTPDNNMT 612
QY 677 -----LFLALVDVAQGPFGHQHDKNVFQLEQ 706
DB 613 WQEWKRVHYLEA---NISQLEQAQIOEKNNYELQK 647

RESULT 14
NME4_MOUSE
ID NME4_MOUSE STANDARD; PRT: 1323 AA.
AC Q03391.
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
GN D-aspartate receptor subtype 2b) (NR2D) (NMDAR2D).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93050214; PubMed=1385220;
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA Inoue Y., Mishina M.;
RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor
channel";
RL FEBS Lett. 313:34-38(1992).
RN [2]
RP REVISIONS.
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,
RA Inoue Y., Mishina M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC -1- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D12822; BAA02254.1; -.
DR PIR; S27224; S27224.
DR HSP; P19491; IGR2.
DR MGD; MGI:95823; Grin2d.
DR InterPro; IPR001320; Ion_glut_receptor.
DR InterPro; IPR001508; NMDA_receptor.
DR InterPro; IPR001311; SBP_glut_receptor.
DR Pfam; PF00060; lig_chan; 1.
DR PRINTS; PR00177; NMDARECEPTOR.
DR SMART; SM00079; PBPe; 1.
KW Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW Ionic channel; Magnesium.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1323 GLUTAMATE [NMDA] RECEPTOR SUBUNIT
FT EPSILON 4.
FT DOMAIN 28 581 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 582 601 1 (POTENTIAL).
FT DOMAIN 602 623 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 624 643 2 (POTENTIAL).
FT DOMAIN 644 653 EXTRACELLULAR (POTENTIAL).
```

```
FT TRANSMEM 654 672 3 (POTENTIAL).
FT DOMAIN 673 841 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 842 861 4 (POTENTIAL).
FT DOMAIN 863 1323 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 278 283 POLY-GLY.
FT DOMAIN 905 913 POLY-PRO.
FT DOMAIN 1030 1035 POLY-ALA.
FT DOMAIN 1197 1201 POLY-PRO.
FT SITE 639 FUNCTIONAL DETERMINANT OF NMDA
RECEPTORS (BY SIMILARITY).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1323 AA; 142907 MW; 8AE9878F90DD0921 CRC64;

Query Match 2.5%; Score 100; DB 1; Length 1323;
Best Local Similarity 22.7%; Pred. No. 30;
Matches 106; Conservative 46; Mismatches 156; Indels 158; Gaps 28;

QY 162 INLAWSGQEAIRVYIALGLTOAEINEFTGPAFLAWGRMGLH-----TWDGGLP 215
DB 48 LNVALVFSGP-----AYAAEAARL-----GPAVAAAVRSPGLDVRPVVALVNGSDP 93
QY 216 PSHHIKQLYLQ-----HRVL--DQMRSEGMTVPVLPAPAGHVPEAVTRVFPQVNT 263
DB 94 -----RSLVLQCLDLLSGLRVHGVFEDDSRAPAVAPILDFLSAQTSLPIVAVHG----- 143
QY 264 KMSGWGHFNCSVCSFLLAPEDPIPIIGSLPLRELIIKEFGTDHIYGADTFNEMQ----- 318
DB 144 -----GAALVLTPEK-----GSTFLQ-----LGSSTEQQLGVIEV 175
QY 319 -PPSEPSYLAATTA-----VYEAMTAVDTEAVMLQGLWFOHQPF-WGPAQIRAVLGA 372
DB 176 LBEYDWTSEVATTPTRAPGHRFLSVIEVLTDG--SLVGW--EHRGALTLDPGAGEAVLGA 231
QY 373 VPRG-----RLVLVDLFAESQPV--TRTASQGGPFITWMLH---NFGNGHILGF-ALE 421
DB 232 QLRVSQAIRLLFC-AREAEAPVFRAAEAGLTGPGYVFMVGPQLGAGGGSGVGEPLL 290
QY 422 AVNGPEARLPNSMTMGVGTGMAPEGISQNEVYSIMAEIAGWKRDPVPLAAVNTSFA-- 479
DB 291 LPGGAPLPAGLF-----AVRSAGWRDRLARVAAGVAVARG 327
QY 480 ----ARRYGVSHPDAGAWRLLLRSVYNGSGEACRHNRSPLVRPSLQ---MNTSIWYN 532
DB 328 AQALLRDYGLF-PEL-----GHDCRAQNRTH--RGESLHRYFMNIT-WDN 368
QY 533 RSDVFEAWRLLTSPSLATSPAFRYDLDLTRQAVQELVSLYEE 578
DB 369 RDSYFNEDGFLVN--PSLV-----VISLTRDRTWEVVSWEQQ 404
```

```
RESULT 15
NME4_HUMAN
ID NME4_HUMAN STANDARD; PRT: 1336 AA.
AC O15399;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl
DE D-aspartate receptor subtype 2b) (NR2D) (NMDAR2D) (EB11).
GN GRIN2D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98149394; PubMed=9489750;
```





Result No.	Query %		Length	DB	ID	Description
	Match	Score				
1	83.9	3306	639	4	Q14769	Q14769 homo sapien
2	83.9	3305	739	11	O54752	O54752 mus musculu
3	83.8	3302	739	11	O88325	O88325 mus musculu
4	66.3	2613.5	753	13	Q90276	Q90276 dromaius no
5	66.3	2611.5	753	13	Q90275	Q90275 dromaius no
6	38.7	1522.5	806	10	Q9FN43	Q9fn43 arabidopsis
7	36.7	1445	811	10	Q9ZRA5	Q9zr45 arabidopsis
8	36.4	1433.5	778	5	Q95TU3	Q9zrtj3 drosophila
9	35.7	1407	690	5	Q9VL15	Q9vll5 drosophila
10	26.9	1060.5	770	16	Q9AAQ6	Q9aaq6 caulobacter
11	24.6	970	575	5	Q9NAP6	Q9na6 caenorhabdi
12	3.6	143.5	1862	2	Q93TW7	Q93tw7 stigmatella
13	3.2	128	444	3	Q96GW3	Q96gw3 gibberella
14	3.2	124.5	783	2	Q9FBV4	Q9fbv4 streptomyce
15	3.1	122.5	3247	12	Q63553	Q65553 bovine herp
16	3.1	121	2116	2	Q93N89	Q93n89 streptomyce

	Matches	617;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	125	TPNRYRYQNVCTQSYSFVWMDWARWEREIDWALMGINLALAWSQEATQWRVYLALGL	184							
		I								
Db	21	TARRYRYQNVCTQSYSFVWMDWARWEREIDWALMGINLALAWSQEATQWRVYLALGL	80							
		I								
QY	185	TOAINEFFTGPALFANGRMGNLHTWDGPLPPSWHIKQLYQHRVLDQMRSGMTVPVLA	244							
		I								
Db	81	TOAINEFFTGPALFANGRMGNLHTWDGPLPPSWHIKQLYQHRVLDQMRSGMTVPVLA	140							
		I								
QY	245	FAGHVPEATRVFPQNVTKMGSGHNCVSCSFLAPEDPIFFIIGSLFLURELIKBFQ	304							

Db	141	FAGHVEAVTRVFPQVNVTKMSGHNFCNSYSCSFLAPEDPIFLIGSLFLRELKKEG	200
QY	305	TDHIYGADTFNEMOPSPSSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOFWGPA	364
Db	201	TDHIYGADTFNEMOPSPSSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOFWGPA	260
QY	365	QIRAVILGAVPRGRLLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGNGHGLFGALEAVN	424
Db	261	QIRAVILGAVPRGRLLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGNGHGLFGALEAVN	320
QY	425	GGPEARLPNSTVMGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYG	484
Db	321	GGPEARLPNSTVMGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYG	380
QY	485	VSPHDAGAARLLRSVYNCSGEACRHNRSPLVRRPSLQMNSTSIWYNSDVFEAWRLLL	544
Db	381	VSPHDAGAARLLRSVYNCSGEACRHNRSPLVRRPSLQMNSTSIWYNSDVFEAWRLLL	440
QY	545	TSAPSLATSPAFRYDLDLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELLP	604
Db	441	TSAPSLATSPAFRYDLDLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAYELLP	500
QY	605	ALDEVLASDSRFLGSLGLEQARAAVSEAEADFEQNSRYQLTLWGPEGNILDYANKOLA	664
Db	501	ALDEVLASDSRFLGSLGLEQARAAVSEAEADFEQNSRYQLTLWGPEGNILDYANKOLA	560
QY	665	GLVANYTTPRWRLFLDALVDSVAGQIPFQOQHFQDKNVFQLEQAFVLSKORYPSQPRGDTV	724
Db	561	GLVANYTTPRWRLFLDALVDSVAGQIPFQOQHFQDKNVFQLEQAFVLSKORYPSQPRGDTV	620
QY	725	DLAKKIFLKYPGWVAGSW 743	
Db	621	DLAKKIFLKYPGWVAGSW 639	
RESULT	2		
ID	054752	PRELIMINARY; PRT; 739 AA.	
AC	054752;		
DT	01-JUN-1998 (TReMBLrel. 06, Created)		
DT	01-JUN-1998 (TReMBLrel. 06, Last sequence update)		
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	NAGLU (EC 3.2.1.50).		
GN	NAGLU.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=B6/CBAFLJ; TISSUE=LIVER;		
RA	Zhao K.W., Li H.H., Neufeld E.F.;		
RT	"Cloning and expression of Mouse gene encoding the lysosomal alpha-N-		
RT	acetylglucosaminidase.";		
RT	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; U85247; AAB8084.1; -;		
DR	MGI; MGI:1351641; Naglu.		
DR	InterPro; IPR002086; Aldehyde_dehydr.		
DR	PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.		
KW	Hydrolase; Glycosidase.		
SQ	SEQUENCE 739 AA; 82610 MW; B414C336A04EF0EF CRC64;		
Query Match	83.9%;	Score 3305;	DB 11; Length 739;
Best Local Similarity	83.0%;	Pred. No. 1.7e-236;	
Matches 611;	Conservative 53;	Mismatches 70;	Indels 2; Gaps 1;
QY	1	MEAVAAVAVGLLLAGAGGAGGAGDEARAAVRLVRLGPGPAANDSVSVERALAAKP	60
Db	1	MEAGLAVILGLFLL--AGGSVGDEAREAKAVRELVRLLGPGPAANFLVSVERALADES	58
QY	61	GLDITSLGGGGAARVVRGSGTGVAAAGLHRYLRDFCCGCHVAVSWSSQLRPLRPVAPVGE	120

Db	59	GLDITSLGGGVPVLRGSGTGVAAAGLHRYLRDFCCGCHVAVSWSSQLRPLRPVAPVPG	118
QY	121	LTEATPNRYRYQNVCTOSYSFVWMDWAREREIDWALNGINLALAWSGEATWQRYVL	180
Db	119	LTEATPNRYRYQNVCTOSYSFVWMDWAREREIDWALNGINLALAWSGEATWQRYVL	178
QY	181	ALGTLQAEINEFFTPAFPAFLWGRGNLHTDGPLPPSWHHIKOLYLQHRVLDQMRSGFWTP	240
Db	179	ALGTLQAEINEFFTPAFPAFLWGRGNLHTDGPLPPSWHHIKOLYLQHRVLDQMRSGFWTP	238
QY	241	VLPFAGHVPFAVTRVFPQVNVTKMSGHNFCNSYSCSFLAPEDPIFLIGSLFLREL	300
Db	239	VLPFAGHVPFAVTRVFPQVNVTKMSGHNFCNSYSCSFLAPEDPIFLIGSLFLREL	298
QY	301	KEFGTDHIYGADTFNEMOPSPSSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOF	360
Db	299	KEFGTDHIYGADTFNEMOPSPSSPSYLAATTAAYEAMTAVDTAEAVMLLQGWLFQHOPOF	358
QY	361	WGPAQIRAVILGAVPRGRLLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGNGHGLFGAL	420
Db	359	WGPAQIRAVILGAVPRGRLLVLDLFAESQPVYTRTASFQGPFTWCMHLNFGNGHGLFGAL	418
QY	421	EAVNGGPEARLPNSTVMGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAA	480
Db	419	EAVNGGPEARLPNSTVMGTGMAPEGISONEVYSLMAELGWRKDPVDPDLAAWVTSFAA	478
QY	481	RRYGSHPDAGAARLLRSVYNCSGEACRHNRSPLVRRPSLQMNSTSIWYNSDVFEAW	540
Db	479	RRYGSHPDAGAARLLRSVYNCSGEACRHNRSPLVRRPSLQMNSTSIWYNSDVFEAW	538
QY	541	RLLTSAFSLATSPAFRYDLDLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAY	600
Db	539	RLLTSAFSLATSPAFRYDLDLTROAVQELVSLYEEARSAYLSKELASLLRAGGVLAY	598
QY	601	ELLPALDEVLASDSRFLGSLGLEQARAAVSEAEADFEQNSRYQLTLWGPEGNILDYAN	660
Db	599	ELLPALDEVLASDSRFLGSLGLEQARAAVSEAEADFEQNSRYQLTLWGPEGNILDYAN	658
QY	661	KQAGLVANYTTPRWRLFLDALVDSVAGQIPFQOQHFQDKNVFQLEQAFVLSKORYPSQPR	720
Db	659	KQAGLVANYTTPRWRLFLDALVDSVAGQIPFQOQHFQDKNVFQLEQAFVLSKORYPSQPR	718
QY	721	GDTVDLAKKIFLKYYP 736	
Db	719	GDTVDLAKKIFLKYYP 734	
RESULT	3		
ID	088325	PRELIMINARY; PRT; 739 AA.	
AC	088325;		
DT	01-NOV-1998 (TReMBLrel. 08, Created)		
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)		
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	ALPHA-N-ACETYLGLUCOSAMINIDASE.		
GN	NAGLU.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129SV;		
RX	MEDLINE=20056274; PubMed=10588735;		
RA	Li H.H., Yu W.H., Rozengurt N., Zhao H.Z., Lyons K.M.,		
RA	Aganostaras S., Fenslow M.S., Suzuki K., Vanier M.F., Neufeld E.F.;		
RT	"Mouse model of Sanfilippo syndrome type B produced by targeted		
RT	disruption of the gene encoding alpha-N-acetylglucosaminidase.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 96:14505-14510(1999).		
DR	EMBL; AF003255; AAC26842.1; -;		
DR	MGI; MGI:1351641; Naglu.		
DR	InterPro; IPR002086; Aldehyde_dehydr.		

DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN.1.  
SQ SEQUENCE 739 AA; 82596 MW; 01C71966557CDE27 CRC64;

Query Match 83.8%; Score 3302; DB 11; Length 739;  
Best Local Similarity 82.9%; Pred. No. 2.8e-236;  
Matches 610; Conservative 54; Mismatches 70; Indels 2; Gaps 1;  
QY 1 MEAAVAAAGVLLLAGAGGAGGADEAAAVRALVRLGPGPAADFSVVERALAAKP 60  
DB 1 MEAAAGLVILGFLLL--AGGSVGDREAKAVRELVRLGLPGPAANFVSVVERALADES 58  
QY 61 GLDTSYSLGGGGAARVRVRSRGSTGVAAGLHRYLRDFCGCHVAVSGSQRLLRPLPVPAGE 120  
DB 59 GLDTSYSLGGGVPVLVRGSGTGAAGLHRYLRDFCGCHVAVSGSQRLLRPLPVPAGE 118  
QY 121 LTEATPNRYRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAWSGOEAIWQRVYL 180  
DB 119 LFEETPNRYRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAWSGOEAIWQRVYL 178  
QY 181 ALGLTQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMSEFGWTP 240  
DB 179 ALGLTQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMSEFGWTP 238  
QY 241 VLPAPAGHVPAVRVFPQVNVTKMGSHFNCSYCSFLLAPEDPIPIIGSLFLREL 300  
DB 239 VLPAPAGHVPAVRVFPQVNVTKMGSHFNCSYCSFLLAPEDPIPIIGSLFLREL 298  
QY 301 KEFGTDHIYGADTFNEMOPPSSEPSYLAATAATVAYEAMVTDPDAVLLQGLWLFQHQPOF 360  
DB 299 KEFGTDHIYGADTFNEMOPPSSEPSYLAATAATVAYEAMVTDPDAVLLQGLWLFQHQPOF 358  
QY 361 WGPQIRAVLAVPRGRLLVLDLFAESQPVYTRTASFGQPFIMCMLHNFEGNHLFGAL 420  
DB 359 WGPQIRAVLAVPRGRLLVLDLFAESQPVYTRTASFGQPFIMCMLHNFEGNHLFGAL 418  
QY 421 EAVNGGPEAARLFNSMTWGTGMAPEGISQNEVYSLMAELGWRKDPVPLAAWVTSFAA 480  
DB 419 EDVNRGQPAARLFNSMTWGTGMAPEGISQNEVYSLMAELGWRKDPVPLAAWVTSFAA 478  
QY 481 RRYGVSHPDAGAAWRLRLRSYVNCSEACRGNHNSPLVRRPQLQMTS IWNRSDFEAW 540  
DB 479 RRYGVSHPDAGAAWRLRLRSYVNCSEACRGNHNSPLVRRPQLQMTS IWNRSDFEAW 538  
QY 541 RLILTSAPSLTSPAFRYDLDTLQRAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 600  
DB 539 RLILTSAPSLTSPAFRYDLDTLQRAVQELVSLYEEARSAYLSKELASLLRAGGVLAY 598  
QY 601 ELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFEQNSRYQLTLWGPENILDYAN 660  
DB 599 KLLPTLDELLASSHFLGTLWLDQARAAVSEAEADFEQNSRYQLTLWGPENILDYAN 658  
QY 661 KQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDFKNVFOLEAFVLSKORYPSQPR 720  
DB 659 KQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDFKNVFOLEAFVLSKORYPSQPR 718  
QY 721 GDTVDLAKKIFLKYYP 736  
DB 719 GDTVDLAKKIFLKYYP 734

RESULT 4  
Q90276 PRELIMINARY; PRT; 753 AA.  
AC Q90276;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LYSSOMAL ALPHA-N-ACETYL GLUCOSAMINIDASE.  
OS Dromaius novae-hollandiae (Emu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Palaeognathae; Casuariiformes; Dromaiidae;  
OC Dromaius.

OX NCBI\_TaxID=8790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21309063; PubMed=11414757;  
RA Aronovich E.L., Johnston J.M., Wang P., Giger U., Whitley C.B.;  
RT "Molecular Basis of Mucopolysaccharidosis Type IIIB in Emu (Dromaius  
novae-hollandiae): An Avian Model of Sanfilippo Syndrome Type B.";  
RL Genomics 74:299-305(2001).  
DR EMBL; AF331668; AAK73654.1; -.  
SQ SEQUENCE 753 AA; 84013 MW; C4B5332A6775EC01 CRC64;

Query Match 66.3%; Score 2613.5; DB 13; Length 753;  
Best Local Similarity 64.7%; Pred. No. 3.3e-185;  
Matches 480; Conservative 105; Mismatches 144; Indels 13; Gaps 6;  
QY 4 VAVAAAGVLLLAGAGG-----AAGDEAREAAVRAVRLGPGPAADFSVVERAL 56  
DB 5 VRLTALAAMAAARAASPPPLPPVAGAEADARQAAVRAALARRLLGPRAA--VALSVDSGL 63  
QY 57 AAKPGDLYSLGG--GGAARVVRGSGTGAAGLHRYLRDFCGCHVAVSGSQRLLRPL 114  
DB 64 AA--GGDLYRVHSPGAAVAVAVAGSGVAAAGLHRYLRDFCGCHLSWSGRQLRLDPL 122  
QY 115 PAVPGELTEATPNRYRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAWSGOEA 174  
DB 123 PRVPAEIRATAPGRFRYQNVCTOSYFVWMDWARWEREDIMWALNGINLALAFAGEAV 182  
QY 175 WQRYLALGALQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMR 234  
DB 183 WQRYLALGALQAEINEFFTPAPFLAWGRMGNLHTWDGPPSWHIKOLYLQHRVLDOMR 242  
QY 235 SFGMTVPVLPAPAGHVPAVRVFPQVNVTKMGSHFNCSYCSFLLAPEDPIPIIGSL 294  
DB 243 SLGMITVLPAPAGHVFPQVNLRAFPVNRATRLGWSHFDCTYSTYLLDLPDPFQVIGTL 302  
QY 295 FLRELKEFGTDHIYGADTFNEMOPPSSEPSYLAATAATVAYEAMVTDPDAVLLQGLWLF 354  
DB 303 FLRELKEFGTDHIYGADTFNEMOPPSSEPSYLAATAATVAYEAMVTDPDAVLLQGLWLF 362  
QY 355 QHQPQWGAQIRAVLGVPRGRLLVLDLFAESQPVYTRTASFGQPFIMCMLHNFEGNH 414  
DB 363 QHQPQWGAQIRAVLGVPRGRLLVLDLFAESQPVYTRTASFGQPFIMCMLHNFEGNH 422  
QY 415 GLFGALEAVNGGPEAARLFNSMTWGTGMAPEGISQNEVYSLMAELGWRKDPVPLAAW 474  
DB 423 GLFGTVEALNHGPFARFPNSMTWGTGMAPEGISQNEVYSLMAELGWRKDPVPLAAW 481  
QY 475 VTSFAARRYGVSHPDAGAAWRLRLRSYVNCSEACRGNHNSPLVRRPQLQMTS IWNRS 534  
DB 482 VARYAERRYGAPNAAASAWQLLRSVYNTG--VCVNHNSPLVRRPQLQMTS IWNRS 540  
QY 535 DYFEARLLTSPATSPAFRYDLDTLQRAVQELVSLYEEARSAYLSKELASLLRA 594  
DB 541 DYFEARLLTSPATSPAFRYDLDTLQRAVQELVSLYEEARSAYLSKELASLLRA 600  
QY 595 GGVLAYELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFEQNSRYQLTLWGPEN 654  
DB 601 GGVLAYELLPALDEVLASDRFLGSLWLEQARAAVSEAEADFEQNSRYQLTLWGPEN 660  
QY 655 ILDYANKQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDFKNVFOLEAFVLSKOR 714  
DB 661 ILDYANKQLAGLVANYTTPRWRLFLDALVDSVAQGIPIFQHQDFKNVFOLEAFVLSKOR 720  
QY 715 YPSQPRGDPVLDLAKKIFLKYYP 736  
DB 721 YPTAPVGDFTLEISKIFLKYYP 742

RESULT 5  
Q90275 PRELIMINARY; PRT; 753 AA.  
ID Q90275  
AC Q90275;







QY	248	HVPEAVTRVFPQVNVTKMGSGWHFNCYSYSCSFLAPEDPIFPFIIGSLFLRELKIEFGTDH	307
Db	267	HVPRALKRLNPSTFMEVQRWQNFQPDYCCGLFVEPTENLKEIASRELFHLNLIITKYSNH	326
QY	308	IYGADTFNEMQPPSSPSYLAATAATVAYEAMTAVTEAVWLLQGLWLFQHQFQFNGPAQIR	367
Db	327	IFFCDPFNELEPPVAKPEYMRSTAAAIYESMRGIDQAIWLLQGMFVKNP-FWTTDMAE	385
QY	368	AVLGAVPRGLLVLDLFAESQPVYTRTASFQGPFTWCMHNFNGHGLFGALEAVNGCP	427
Db	386	AFTAAPRGRILVDLQSQQFQOYELTRSYFGQPIWCMHNFNGTGLMFGSAKLNSGI	445
QY	428	EAARLPNSTWGTGMAPEGISQNEVYVYSLMAELGWRKDPVDPDLAAWYTSFAARYGVSH	487
Db	446	EEARRLPNSSLVGTITPEGIGQYVMSFTLGRGWSNTSL-DLDSWETNFSHSRYGKD	504
QY	488	PDAGAARLLRLSRVYNCSG-EACRGHNRSPLYRRSLQWNTSIWNRSDVFEAWRLLLTS	546
Db	505	ERLEQAWLLKNSVYFRGLQKNRG--QVYVTRRPSFNQPTWYTNASAVLDHLLITL	562
QY	547	APSLATS---PAFRYDLDLTRQAAQVELSVLYEARSAYLSKELASLLRAGGYLAYEL	602
Db	563	RAIPLDNRNVEIYEDLVDTITRFQIISADQLYINLRSAYRKRQVSRF---EFLSVKL	618
QY	603	LPALDE---VLASDRFLGSLGWLEQARAANVSEAEADFYEQNSRQTLTWGPEGNILDYA	659
Db	619	LKLFDDMELTASRNFLLGNWLQQAKAAPNTGQORNEEFNARNOITAWGPDGQLLDYA	678
QY	660	NKQLAGLVANVYPRWRLFLFALVDSVAOCIPQOQHQFDKNY-FQLEQAFVLSKORYPSQ	718
Db	679	CKQWSGLVSDYPRWRRLFLEDYTVLHAGRPNGTAFKLKYSHEITELPFSNKDDVYPT	738
QY	719	PRGDTVDLAKKIF 731	
Db	739	PVGNTWLISQDIF 751	
RESULT	9		
Q9VLL5			
ID	Q9VLL5	PRELIMINARY;	690 AA.
AC	Q9VLL5;		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE	CG13397 PROTEIN		
GN	ESTS:172F5T OR CG13397		
OS	Drosophila melanogaster (Fruit fly)		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
EC	Oxytrichoidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RP	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=107196006; PubMed=10731132;		
RA	Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ball R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Bescon K.Y., Bencs P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Doudou K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.		
RA	Fosler C., Gabriellian A.E., Garg N.C., Gelbart W.M., Glasser K.,		
RA	Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		

Query Match	35.7%	Score 1407	DB 5	Length 690
Best Local Similarity	43.3%	Pred. No. 8.6e-96		
Matches 290	Conservative 102	Mismatches 255	Indels 22	Gaps 11
QY 74	RVVRGSGTGVAAAGLHRYLRDRCGCHVAVGSGQLRPR--PLPVPAGELTEATPNRYRY	131		
DB 6	RILLMGWDGVSCKALHLYLVNKKVDWFKMRLELTPTNLQLPNWTIESKASP--IY 63			
QY 132	YONVCTOSYSFVWDDWARWEIDMMALNGINLALAWSQEAIAWQVRYLALGLQAEINE	191		
DB 64	HQNVCTWSYSFAWGWIEQWRHLDHMLGMLSLTIA-PVQEAIVKVVYTDMLRMEIDE	122		
QY 192	FPTGPAPLAWGRMGNLHTWDGGLPSPWHIKQLYLQHRVLDQWRSGMTPVLPFAFAGHVP	251		
DB 123	HLAGPAPQAWQRMGNIRGWAGPLTPAWRYYQLLQOEIITAQRNLGMSVALPFAFAGHVP	182		
QY 252	AVTRFEPQVNVTKMGSWGHFNCYSYSCSFLAPEDPTFPIIGLSFLRELKEFGTHIYGA	311		
DB 193	ALKRLNPSTFMEVQRWQNFDPDRYCCGLFVPEITENLFKEIASRFLNIIKTYGSHNIFPC	242		
QY 312	PTFENMQPPSSPSYLAATTAAYGAMTAVDEANVWLLQGLWFQHQPFQWGPQAQIRAVLG	371		
DB 243	DPFNELEPPVAKPEYMRSTAAAIYSEMRGIDPQAIWLLQGMFVKNP--EWTDMAEAFLT	301		
QY 372	AVPRGLLVLDFAESQPVYTTASFGQPFVTCMLHNFEGNGHGLFGALAEVANGGPEAAR	431		
DB 302	AAPRGILVLDLQSQFPQYELTRSYFGQPFVTCMLHNFEGTLGMEGSAKLINGIEEAR	361		
QY 432	LPFNSTMVGTGMAGEISQNEVYVYSLMAELGNWRKDDVPDLANWVTSFAARRGVSHPDAG	491		
DB 362	RLPNSSLVGTGITTEPQIGQVYVYSPFLTRGWSNTSL-DLDSWFTTFSHRYGVKVDLER	420		
QY 492	AAWRLLLRVYVNCSG-EACRGHNRSLPLVRPSLQMTSTIWNYSVDVEAWRLLLLTSPSL	550		
DB 421	QAWLLLNKNSVYFSGQLQKWRG--QYVTVTRPSFNQPFVYVNASAVLDWHLLLTFRAL	478		
QY 551	ATS----PAFRYDLDLTRQAVQELVSLYEEARSAYLSKELASLRAGVLALEYLLPAL	606		
DB 479	PLEDRNYEYHDLVDITRQFLQISADQLYINLRSAYRKRQVSRF----EFLSVKLLKLF	534		
QY 607	DE---VLASDRFLGSLWEQARAAVSEAEADYEQNSRYQLTLGPGENILDYANKQL	663		
DB 535	DDWELTLASRNFLLGNWLOQAQAPNTGQQRNFENFARNQITANGPDQQLIDYACKQW	594		
QY 664	AGLVANVYTPRWRLEALVDSVAQGIQFQHQFDKNV-FQLEQAFVYLSKORYPSOPRGD	722		
DB 595	SGLVSDYIPRWRLEFLEDVTVLHAGRPNGTAFKLKVSHEILPFSNKKDDVYVPTVGN	654		

```
QY 723 TVDLAKKIF 731
D 1 : : : :
Db 655 TWLISQDIF 663

RESULT 10
Q9AAQ6 PRELIMINARY; PRT; 770 AA.
AC Q9AAQ6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ALPHA-N-ACETYLGALUCOSAMINIDASE.
GN CC0540.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005727; AAK22527.1; -.
DR TIGR; CC0540; -.
KW Complete proteome.
SQ SEQUENCE 770 AA; 84287 MW; 94A694F41C18DB20 CRC64;

Query Match 26.9%; Score 1060.5; DB 16; Length 770;
Best Local Similarity 33.0%; Pred. No. 5e-70;
Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;

QY 2 EAVAVAAVGVLLLAGAGGAAGDEAREAAVVALVRLGPGPAADFVSVERALAAKPG 61
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 QAISLAAAFVSPALAAAGSTG- - - - -VAAARASLKRUFGRRLAG- - - - -AHLTVTPG 59

QY 62 LDT- - -YSLGGGAARVVRGSGTGVAAAAGLHRYLRDFCGCHVAVSGSOLRLPRPLPAVPG 119
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 AERSWYAIGKGGA-ISISGSPVALVRGAYAHLRQAGLAHVSWEGDRVQVQAGVAPAG 118

QY 120 ELTEATPNRYRYQVQTSYSEFVVDWAREREIDMMALNGINLALAWSGOEAIWQVY 179
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 ARVE-TPFRHAYLNTCTGYTTPWGMGRWTRIDWMAAHGIDMPLAMEGQEVVRAIW 177

QY 180 LALGLTQAEINEFFTGPAFLANGMNLHTWDGPPSHIKQLYLRVLDQMSFGMT 239
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 REFLSEALADYFGSPAFTPHRMGNTGYKAPLPTAWIDKKDLQVKILGRMSLGMT 237

QY 240 PVLPAFAGHVAEATRVFPQVNVTKMGSHGFCNSCSFLIAPEDPIFFPIIGSLFLREL 299
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 PTLPAFGVYVPAFAEKNPKARIYRMWPWEGFHET- - - - -WLDPAADPLFAKIAARFLALY 293

QY 300 IKEFGTDHYGADTNEWQPP- - - - -SSEPSY- - - - -LA 328
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 TETFGAGTYLADSFNEMLPPINADGADARDAAYGTANTVTKVEVDPAKKAORLA 353

QY 329 AATTAYEAMTAVDTEAVMLLQGLWLFQHPQFQWGAQVAVGPRGLVLDLFAESQ 388
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 AVGKAIYDSIRQTRDAVWVWQGLFGADSHFWDPAISLYSLVDDKMLILDIGNDY 413

QY 389 P-VYTRTASFQOPTICMLNFGNGHGLFGAL- - - - -EAVNGGPEARLFFPNSTMVG 440
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 PNWKNAKAFGKGIYGVVHNYGGSNPVYGDLYRQDIPAIANPDAGKL- - - - -AG 467
```

```
QY 441 TGMAGEISQNEVYVYSLMAELGWRKDPVDPDLAAWVTSFAARYGVSHPDAGAAWRLLR 500
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 FGMFPEGLHNNSIVYEAVYDLAWSEQASP-ATWLTTRYARARYGKTSPLDALGOLVPA 526

QY 501 VVNCSEACRGHNRSP- - - - -LVRRPSLQWNTSIWY- - - - -NRSDFEAWRLLLTSAPSLS 554
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 527 AFSTRYWSPRWMSKAGAYLFFKRPRTATYGFDPFHQPDRAKLEAAVKALTALPTYGQSP 586

QY 555 AFRYDLDLTLTROAVQELVSLYYEEARSAYLSKELASLLRAGGVLAVELLPALDEVLASD 614
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 587 LFVLDLTDATRLATMKIDDLQVAAVYRGDTA- - - - -AGDAARVEI- - - - -EALALSI 637

QY 615 RELLG- - - - -SWLEQARAAVSEAEADYEQNSRYQLTLWPGEGNILDYANKQLAGLV 667
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 638 DKLLGVPQDTLATWIDEARAYGDTPADAAAYVANAKAQTWNGEGNLDYASKAWQGLY 697

QY 668 ANYYTPRWLFLEALVDSVAGQIPFQOHQFDKNVQLQEQAFVLSKQRY- - - - -PSOPRGDT 723
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 698 KSFYLPWRSRFLDAL- - - - -KAAGTGTFDEVTVTRGGVAVERAWVEAYVRRKPADPIGEI 755

QY 724 VDLAKKI 730
D : : : :
Db 756 KTLIARI 762

RESULT 11
Q9NAP6 PRELIMINARY; PRT; 575 AA.
AC Q9NAP6;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE K09E4.4 PROTEIN.
GN K09E4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1];
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2];
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z83234; CAB70170.1; -.
SQ SEQUENCE 575 AA; 68849 MW; FE7216DB39F4D93C CRC64;

Query Match 24.6%; Score 970; DB 5; Length 575;
Best Local Similarity 35.7%; Pred. No. 1.7e-63;
Matches 210; Conservative 104; Mismatches 248; Indels 26; Gaps 10;

QY 135 VCTOSYSEFVVDWAREREIDMMALNGINLALAWSGOEAIWQVYVYVYVYVYVYVYVYVY 194
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MCTSYSEFVVDWAREREIDMMALNGINLALAWSGOEAIWQVYVYVYVYVYVYVYVYVY 60

QY 195 GPAFLANGMNLHTWDGPPSHIKQLYLRVLDQMSFGMTVPLPAFAGHVAEATRV 254
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 SQALVHWRMGNLKYAGGGLSDAQLNDHNLAKRIIDRLLELGTPLPTFAGVDPHLE 120

QY 255 RVFPQVNVTKMGSHGFCNSCSFLIAPEDPIFFPIIGSLFLRELKEFGTD- - - - -HIYGAD 312
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 TLFPAFASKNRLPRNNFTSETSCMLSVSPDFLQFKIGSTFLRHQKMFQGDVNMYSAD 180

QY 313 TFNEMQPPSS- - - - -EPSYLAATTAATVYEAATVDEAVWLLQGLWLFQHPQFQWGAQV 369
D : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 PFNEILPSESAKFAKFEVVKQTAQAINMSCKKVDKNCVWVLLQSWSFYTDQ- - - - -WPAWAIKSF 238
```

```
QY 370 ICAVPRGRLVLDLFAESOPVYTRTASFOGQPFIMCLHNEFGNHGLFGALEAVNGGPEA 429
Db 239 LSAIPVGNLLIIDLVAEPVQWMTSSFOGHFVWCLLHNEFGGSELKGNLQKIDKGVL 298
QY 430 ARLFNPSTWGTGMAPEGISQNEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYGVSHPD 489
Db 299 ALMKAGSNLVGAGLSNEAIDQNVYVQFMIDRMWSEPLP-LNNMLKAYSESRYSADEFK 357
QY 490 AGAAMPLLRSVYNGSGEACRGNHNSPLVRPSLOMNTSIWYNSRDVFEAWRLLLTS-AP 548
Db 358 AQKFWTLTACTYNOPEKWTGTPRESVFLYHREFGCRKIEYFPVEETFSRFRRELLPALV 417
QY 549 SLATSPAFRYDLDLTROAVQELVSLVYEEARSAYLSKELASLL-----RAGGVLAYELP 604
Db 418 VLGEHPLFREDLNDVNMETQ-----FEMNGEALNSSEAFELMEDKQOVGA--SCLEML 469
QY 605 ALDEVLASDRFLGSLWLEQARAAVSEAEADFYENSRVQLTLWGPEGNIIDYANKQLA 664
Db 470 EMFOKLESYNSRDVROWIENAKSIAPTSEERQVFPVPTAGDILTVMGPTGNLDYAHREWA 529
QY 665 GLVANYTTPRWRLFLEALVDSVAOGIPFOQHOHFDKNVFO-LEQAFVLS 711
Db 530 GLMSGYGRRWQYF-----CDWILEHDFNHTFESISVFRDVERPFSIS 573

RESULT 12
Q93TW7 PRELIMINARY; PRT; 1862 AA.
AC Q93TW7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MXAE.
GN MXAE.
OS Stigmatella aurantiaca.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
OX NCBI_TaxID=41;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21110452; PubMed=11182319;
RA Sliakowski B., Nordsiek G., Kunze B., Blocker H., Muller R.;
RT "Novel features in a combined polyketide synthase/non-ribosomal
RT peptide synthetase: the myxalamid biosynthetic gene cluster of the
RT myxobacterium Stigmatella aurantiaca Sgal5."
RL Chem. Biol. 8:59-69(2001).
DR EMBL: AF319998; AAK57189.1; -.
SQ SEQUENCE 1862 AA; 200005 MW; E809DCEC6A320144 CRC64;

Query Match 3.6%; Score 143.5; DB 2; Length 1862;
Best Local Similarity 22.3%; Pred. No. 0.19;
Matches 180; Conservative 79; Mismatches 289; Indels 259; Gaps 37;

QY 3 AVAAVAVGVLLLAGAGGAAGDEAREAAVVALVRLGPGPAADFSVVERALAAKPG 62
Db 962 AVPASALVELGLSAAAEALGAGARELSDVEFERALVLTADRLVGVHLSPASGGQHV 1021
QY 63 DTVSLGGGAA-----RVRVRSSTGVAAAAGLHRYLRDFCGCHVAVSGQLRLPR- 112
Db 1022 HHSRAVGGTSPGAGWVHRCGQIRAAATPSGSPSEVLDAVRSQAQHVQGAAYVEQLERC 1081
QY 113 -----PLPAV-----PGEUTEATPNRYRYQNVCTQSYSFVWMDWARERE----IDWM 157
Db 1082 NVQYEAPLRLTGEARRRPGALGAVALSPEVQ-----ESARYQLHPALD-A 1128
QY 158 ALNGINLALASGOEAIWQRYLALGLTQAEINEFFTGPAFLWG-----RMG 205
Db 1129 GLQTLALALAAESGEAV---LFMPLSTGSLECVQ---GRADVKAHVSIANAATSPEDRMG 1182
QY 206 NLHTWDG-----PLPPSHWIKQLYLOH---RVLDQMS-----FGMTPLVPFAGHVP 250
```

```
Db 1183 TLELDGEGRRVAVARGVRLRRVAAERLLEVLGSEARSOEWFHDVWEPRAVTAQPGPAD 1242
QY 251 EAVTRVFPQVNVTKMSGHF-----NCYSYSCSFLLAPE-----DPIFFPIG 292
Db 1243 WLV-----FLDRGGWGTALVEEIGEQGQPCQVTVTAGETFFQDARFRVNVNKRPEDM 1294
QY 293 SLFLRELI-----KEFGDHYHIGADTFNEMQPPSPSEPSYLAANATVAYEAM----- 338
Db 1295 ERLRLPALPAGHEGRAVYLVGLDAVLDEQGTPESS--VAALHLVKALMGSPARALW 1352
QY 339 -----TAVDTAVWLLOGWLF-----QHOPOEWGPAQIRAVLGA VPRGRLLVLIDL 383
Db 1353 VYTRGAQVTGCTERVSLAQAPLGMGSRVSLQPCVWGG-----LIDL 1396
QY 384 FAESOPVYTRTASFOGQPFIMCLHNEFGNHGLFGALE-----AVNGGPEAR 431
Db 1397 APER--VAEETAALVRE-----ISAFGGDGEDQIALREKRSVPRIARGRVNAPAEPLR 1448
QY 432 LPPNSTWGTGMAPEGISQNEVYSLMAELGWRKDPVDPDLAAWVTSFAARRYGVSHPDAG 491
Db 1449 LRPDGAYLVTTGG-----LGLGLK-----VARWLVARGARHL----- 1480
QY 492 AAWRLLRVYNCSGEACRGNHNSPLVRPSLOMNTSIWYNSRDVFEAWRLLLTSAPLSA 551
Db 1481 ----VLLGRSGASGAGDAASARREGI-----ESLRALGASVTTLA 1517
QY 552 TSPAFRYDLDLTROAVQELVSLVYEEARSAYLSKELASLLRAGGVLAYELLPALDEVLA 611
Db 1518 VDVADREKLAALLREAAATL-----PPLRGVIHAAALLTESLENMD--LA 1561
QY 612 SDSREL----LGSWL--EQARAAVSEAEADFYENSRVQLTLWGPEGNIIDYANKQLAG 665
Db 1562 AMTAMMRPKVLGSLWHLHEVTR-----EALDFFVWFSSSTS-TLWGASG-LAHYA----AG 1610
QY 666 LVANYTTPRWRLFLEALV-DSVAOGIP 691
Db 1611 -----NQFLERLAHHRRAQGLP 1627

RESULT 13
Q96W93 PRELIMINARY; PRT; 444 AA.
AC Q96W93;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TR18.
GN TR18.
OS Gibberella zeae (Fusarium graminearum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5518;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=88-1;
RX MEDLINE=21318682; PubMed=11425709;
RA Lee T., Oh D.W., Kim H.S., Lee J., Kim Y.H., Yun S.H., Lee Y.W.;
RT "Identification of Deoxyvalenol- and Nivalenol-Producing Chemotypes
RT of Gibberella zeae by Using PCR."
RL Appl. Environ. Microbiol. 67:2966-2972(2001).
DR EMBL: AF336365; AAK53574.1; -.
SQ SEQUENCE 444 AA; 47659 MW; BC5208C72484C701 CRC64;
```

Query Match 3.2%; Score 128; DB 3; Length 444;  
Best Local Similarity 20.3%; Pred. No. 0.32;  
Matches 97; Conservative 54; Mismatches 160; Indels 166; Gaps 24;

```
QY 157 MALNGINLALA-WSGOEAIWQRYLALGLTQAEINEFFTGPAFLWGRMGNLHTWDGPLP 215
Db 1 MALDRLLFLLSLWLG-----FVGAQAALSE-----PLP 29
QY 216 PS---WHIKOLYLOHRLVDQMSRSGMTPLVPAPAGHVPVTRFPQV-NVTXMSGNGHF 271
```



```
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwyzer M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=JURA;
RX MEDLINE=971164286; PubMed=9010999;
RA Schwyzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA LaBoissiere S., Misra V., Vleck C., Paces V.;
RT "Gene contents in a 31-kb segment at the left genome end of bovine
RL herpesvirus-1.";
RN Vet. Microbiol. 53:67-77(1996).
RP [5]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyzer M., Vleck C., Lowery D.E., Bello L.J., Meyer G., Misra V.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyzer M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z78205; CAB01605.1; -.
DR EMBL; AJ004801; CA006097.1; -.
SQ SEQUENCE 3247 AA; 332190 MW; 3AEAA72F8F001F6A CRC64;

Query Match      3.1%; Score 122.5; DB 12; Length 3247;
Best Local Similarity 22.9%; Pred. No. 15;
Matches 192; Conservative 68; Mismatches 297; Indels 283; Gaps 45;

QY 1 MEAVAVAAVGVLLLAGAGGAAGDEAREA--AAVRAL--VAR---LIGPGPAADFSVSVE 53
DB 873 MSAARAAARAAGA--AAEAPGAGAEADAEEAAVVGADVDLRQLDALATCAAADREARA- 929
QY 54 RALAAKPLGDTYSLGGGGAARVRVGTGVAANAAGLHRY-----LRDFCGCHVWMSGQ 107
DB 930 RVAAARAGLEAQAAL---AARRRY-AELGERRAELYQRLDALLPLQGFAG----- 975
QY 108 LRLPRPLPVPGETEATPNRYRYQNVCTQSYFVWMDWAREREIDMMLNGINLALA 167
DB 976 -----LRAAPGAL-----ERLCADAGAQSAEDCARFLOEAPPVRAAGV----- 1013
QY 168 WSGQEAIRQRYVYALGLTQAEINEFFTGPAFLAWGRMGNLHTWDGPLPPSHIKQLYLQH 227
DB 1014 --QDRLWOLF-----GRYREALNPSSALAPGALAGL---GPAAFA----- 1048
QY 228 RVLQDMRSFGWTPVLPA-FAGH---VPEAVTRVFPQVNVTKMSGWHGFCNSYCSFLLAP 283
DB 1049 -VLGTALQAMGPAARAFFECHAERVAARAAVAARAAE-----P 1084
QY 284 EDPIFFPIIGSLFLRELKEF--GTDHI-----YGDATFENEMQPP 320
DB 1085 AAPAAVAAAASALREAAEAELRAGGAHVPPPEPAFLDALQAQYAAARREAQTGARRLGAALAA 1144
QY 321 SSEPS-VLAAATTAAYEAMT-----AVDTEAWLLQGMFLQHQPOFGWPAQIRAVLGAVPR 375
DB 1145 VSEASGALASAASALKNAWTAGADAETA-----AAAAADAALGAEE 1188
QY 376 GRLLVLDFLFAESQPVYTTASFGQPFICWMLHNFQGNHGLFGALEAVNGGPEAARLPFN 435
DB 1189 ----VLRAADEA-----AVAAGGE-----DADGGEDASEADAGDAGDAPPAGRLGPQ 1231
QY 436 STMVCTG-----MAPEGISQNEVYSLMAELGWRKD----- 466
DB 1232 AAKLHADLRKARALARRRAEEVRRLRAEERRAEGAAARQOEERWRDLRAALERVEANS 1291
QY 467 --PVPDLAAWYTSFAARRYGVSHPDAGAAWRLLLRSYVNCSG--EACRGNH-RSPL----- 517
DB 1292 AFDAAEALARLAAAAAARGCY-PRPLARQADRALEASAQRATAATEAVLAFNPHAPENAKI 1350
QY 518 -VRRPSLQWNTSIWYN-----RSDVFEAWRL-----LLTSAPSLATSPAF 556
```

```
DB 1351 DVPPPPLAQLRGIAWMDAFSLAAPALSALFPAAADVGLTSLMHIATGLLTFAASGGEPRY 1410
QY 557 RYDLL-----DLTROAVQELVS--LYYEEARSAYLSKELASLLRAGGVLAYELLPALDEV 609
DB 1411 -YDAVTYLEGDLA--AVPLAKYVAFYRRGHAHF-EEERARL-----GALRADVLQAAG-- 1460
QY 610 LASDSRFLLGSLWEQARAAAVSEAEADFYEQNSRYQLTLWGPPEGNTILDYANKQLAGLVAN 669
DB 1461 -----GRAAEISRAR-----RGNVLRSPAERRA-LEAG 1487
QY 670 YTPRWRRLFEALVDS-VAQGIPFQOHQDKNVFQLEQAFVLSKQRYPSQPRGDTVDLAK 728
DB 1488 V---RLELPSEALIASALAQLERFDHARFAGSAY---EAQMLSTVRRDAAAAAREALEAAK 1541

Search completed: September 21, 2002, 12:29:08
Job time: 278 sec
```